InQuest Extended Technical Report

Matthew Russo[†], Tatsunori Hashimoto[†], Daniel Kang[*], Yi Sun[‡], Matei Zaharia[†]
Stanford University[†], University of Illinois Urbana-Champaign[*], University of Chicago[‡]

December 2022

Abstract

Given a streaming dataset $\mathcal{D}$ and an aggregation query, InQuest performs stratified sampling over the stream and uses oracle predictions on the sampled frames to compute a query estimate. InQuest uses a tumbling window to process the stream in "segments". In each segment InQuest draws $N = N_1 + N_2$ samples, where N_1 denotes a number of defensive samples and N_2 denotes a number of samples dynamically allocated to the strata. We show that the mean squared error of InQuest's estimate converges to zero at a rate of $O\left(\frac{1}{N_1} + \frac{N_1}{N_2^2} + \frac{1}{N_2\sqrt{N_1}} + \frac{1}{N_2\sqrt{N_1}t}\right)$ on stationary streams. We further show that InQuest's sample allocation converges to the optimal sample allocation on such streams at a rate of $O\left(\frac{1}{N_1(t-1)}\right)$.

Symbol	Description
$\mathcal{D}$	Streaming dataset of records
$\mathcal{S}$	Stratification, i.e., k strata
$\mathcal{P}(x)$	Proxy model
T	Number of segments (including pilot segment)
N	Per-segment user-specified sampling budget
N_1	Per-segment defensive sample budget
N_2	Per-segment dynamic sample budget
N_{pilot}	Pilot sampling budget
K	Number of strata
$\mathcal{O}(x)$	Oracle predicate
$\mathcal{D}_{tk}$	The set of dataset records in segment T and stratum k
$X_{tk,i}$	ith sample from $\mathcal{D}_{tk}$
X_{tk}	The set of samples drawn from $\mathcal{D}_{tk}$
X_{tk}^+	The set of predicate matching samples drawn from $\mathcal{D}_{tk}$
p_{tk}	Predicate positive rate
w_{tk}	$\lvert D_{tk}\rvert p_{tk}/\sum \lvert D_{tj}\rvert p_{tj}$
σ_{tk}	the true std. dev. of the samples in $\mathcal{D}_{tk}$
a_{tk}^*	the optimal fraction of N_2 allocated to $\mathcal{D}_{tk}$
$f(x)$	statistic function

Table 1: Summary of notation.

1 Setup

Let $\mathcal{D} = \{x_i\}$ be a streaming dataset of records and let $f(x_i) : \mathcal{D} -> \mathbb{R}$ be a statistic function. Furthermore, let $\mathcal{O}(x)$ be the predicate function where $\mathcal{O}(x) \in \{0, 1\}$ for queries with a predicate and $\mathcal{O}(x) \in \{1\}$ for queries without one. We denote $\mathcal{D}^+$ to be the subset of $\mathcal{D}$ that satisfies the predicate function. We wish to compute an estimate of the value $\mu = \mathbb{E}_{x\sim D^+}[f(x)]$. Our objective is to minimize the mean squared error

(MSE) between our estimate $\hat{\mu}$ and the groundtruth value: $\mathbb{E}[(\hat{\mu} - \mu)^2]$.

Refer to tab:notation for a summary of the notation used throughout this document. To compute our query estimate InQuest leverages a combination of stratified and defensive sampling. InQuest uses a tumbling window to divide the stream into "segments". The stream is further split into K disjoint strata based on the per-record proxy scores. Given the true means μ_{tk} and weights w_{tk} for each segment and stratum pair $\mathcal{D}_{tk}$, the query result μ can be computed as:

$$\mu = \sum_t \sum_k w_{tk} \cdot \mu_{tk} \tag{1}$$

In this analysis we assume that $X_{tk,i}$ is a sub-Gaussian random variable with nonzero standard deviation. This enables us to upper bound functions that sum sub-Gaussian variables (e.g., μ_{tk} and σ_{tk}^2) with constants such as $C^{\mu_{tk}}$ and $C^{\sigma_{tk}^2}$. We further assume that at least one stratum has non-zero p_{tk}. Additionally in Sections 5 and 6 we assume that the stream is stationary. Specifically, we assume:

$$\sigma_{tk} = \sigma_{rk} : \forall t, r \in T \tag{2}$$
$$p_{tk} = p_{rk} : \forall t, r \in T \tag{3}$$
$$w_{tk} = w_{rk} : \forall t, r \in T \tag{4}$$

The rest of this document proceeds as follows. In Section 2 we will give a high-level overview of the InQuest algorithm. In Section 3 we will compute the optimal expected error assuming perfect knowledge of μ_{tk}, σ_{tk}, and p_{tk}, before using this result to compute the optimal sample allocation in Section 4. Then in Section 5 we will show that InQuest's sample allocation converges to the optimal sample allocation on stationary streams. In Section 6 we will show that the expected error of InQuest's estimate(s) converges to zero at the specified rate. Finally, we provide an appendix and references in Sections 7 and 8.

2 Algorithm

Algorithm 1 Pseudocode for InQuest. See full paper for definitions of subroutines.

function InQuestPilot($\mathcal{D}_1$, N_{pilot}, K)
  $X_1 \leftarrow \text{UniformSampling}(D_1, N_{\text{pilot}})$
  $X_1^+ \leftarrow \{x | x \in X_1, \mathcal{O}(x) = 1\}$
  return X_1, X_1^+

function InQuest($\mathcal{D}$, $\mathcal{O}$, $\mathcal{P}$, K, N_1, N_2)
  $X_1, X_1^+ \leftarrow \text{InQuestPilot}(\mathcal{D}_1, N_{\text{pilot}}, K)$
  for $t \in [2, 3, \ldots]$ **do**
    $\hat{\mathcal{S}}_t \leftarrow \text{GetStrata}(\mathcal{P}, \mathcal{D}_{t-1}, K, \alpha, \mathcal{S}_{<t})$
    $\hat{a}_t \leftarrow \text{GetAlloc}(\mathcal{D}_{t-1}, K, N_1, N_2, a_{<t}, X_{t-1}, X_{t-1}^+)$
    $\mathcal{D}_{t1}, \ldots, \mathcal{D}_{tK} \leftarrow \text{SplitStream}(\mathcal{D}_t, \hat{\mathcal{S}}_t)$
    for $k \in [1, \ldots, K]$ **do**
      $X_t \leftarrow \text{ReservoirSampling}(\mathcal{D}_{tk}, N\hat{a}_{tk})$
      $X_t^+ \leftarrow \{x | x \in X_t, \mathcal{O}(x) = 1\}$
  $\hat{\mu} \leftarrow \text{GetPrediction}(X_1, \ldots, X_T, X_1^+, \ldots, X_T^+, \mathcal{D})$
  return $\hat{\mu}$

The pseudocode for InQuest is shown in alg:algo. InQuest splits the stream into segments using a tumbling window and uses the first segment to perform pilot sampling. InQuest uses the pilot samples to stratify the stream records by proxy score into K disjoint strata $\mathcal{S}_1, \ldots, \mathcal{S}_K$. InQuest applies N_1/K defensive samples to each stratum and dynamically allocates the remaining N_2 samples across the strata. At any point in time, the the query estimate $\hat{\mu}$ can be computed as a weighted average of the oracle predictions on the frames sampled so far. InQuest computes this estimate as follows:

$$\hat{\mu}_t = \sum_{k=1}^{K} w_{tk} \cdot \frac{\sum_{x \in X_{tk}^+} f(x)}{|X_{tk}^+|} \tag{5}$$

3 Optimal Expected Error

In this section, let us assume that we are given a query with a predicate. Additionally, let us assume that the sample allocations a_{tk} are fixed and given, and let us assume that all p_{tk} are known.

Theorem 1. Given the assumptions stated above, the expected error of our estimator $\hat{\mu}_t$ is:

$$\mathbb{E}[(\hat{\mu}_t - \mu_t)^2] = \sum_{k=1}^{K} \frac{w_{tk}^2 \sigma_{tk}^2}{p_{tk}\left(\frac{N_1}{K} + N_2 a_{tk}\right)} \tag{6}$$

Proof. We define $\mathbb{E}_{x \sim \mathcal{D}_t^+}[f(x)|x \in \mathcal{D}_t^+] = \mu_t$. We wish to compute the error $\mathbb{E}[(\hat{\mu}_t - \mu_t)^2]$ of our estimator $\hat{\mu}_t$ under the assumptions stated above. Our estimator is defined by the following:

$$\hat{\mu}_t = \sum_{k=1}^{K} w_{tk} \cdot \frac{\sum_{x \in X_{tk}^+} f(x)}{|X_{tk}^+|} \tag{7}$$

We'll proceed by decomposing our error into a bias and a variance term and computing each one individually:

$$\mathbb{E}[(\hat{\mu}_t - \mu_t)^2] = \mathbb{E}[\hat{\mu}_t - \mu_t]^2 + Var[\hat{\mu}_t] \tag{8}$$

We will use two lemmas to derive each error term in the decomposition. First we will show that our estimator is unbiased. Then we will derive the error for the variance term.

Lemma 1. The estimator $\hat{\mu}_t$ is unbiased, therefore $\mathbb{E}[\hat{\mu}_t - \mu_t]^2 = 0$.

Proof. We wish to show that our estimator is unbiased – i.e. that $\mathbb{E}_{x \sim \mathcal{D}_t^+}[\hat{\mu}_t] = \mu_t$. We'll start by rewriting μ_t in terms of $\mu_{t1}, \ldots, \mu_{tK}$, and then we'll try to show that $\mathbb{E}_{x \sim \mathcal{D}_t^+}[\hat{\mu}_t]$ simplifies to this rewritten expression.

From our definition of μ_t we have that:

$$\mu_t = \mathbb{E}_{x \sim \mathcal{D}_t^+}[f(x)|x \in \mathcal{D}_t^+] = \sum_{x \in \mathcal{D}_t^+} \frac{1}{|\mathcal{D}_t^+|} \cdot f(x) \tag{9}$$

$$= \frac{1}{|\mathcal{D}_t^+|} \sum_{x \in \mathcal{D}_t^+} f(x) \tag{10}$$

Because $\mathcal{D}_{t1}^+, \ldots, \mathcal{D}_{tK}^+$ are disjoint and span the entirety of $\mathcal{D}_t^+$, we can rewrite this as:

$$\mu_t = \frac{1}{|\mathcal{D}_t^+|}\left(\sum_{x\in\mathcal{D}_{t1}^+} f(x) + \cdots + \sum_{x\in\mathcal{D}_{tK}^+} f(x)\right) \tag{11}$$

$$= \frac{1}{|\mathcal{D}_t^+|}\left(\frac{|\mathcal{D}_{t1}^+|}{|\mathcal{D}_{t1}^+|}\sum_{x\in\mathcal{D}_{t1}^+} f(x) + \cdots + \frac{|\mathcal{D}_{tK}^+|}{|\mathcal{D}_{tK}^+|}\sum_{x\in\mathcal{D}_{tK}^+} f(x)\right) \tag{12}$$

$$= \frac{1}{|\mathcal{D}_t^+|}\left(|\mathcal{D}_{t1}^+|\cdot\mu_{t1} + \cdots + |\mathcal{D}_{tK}^+|\cdot\mu_{tK}\right) \tag{13}$$

$$= \frac{1}{|\mathcal{D}_t^+|}\sum_{k=1}^{K} |\mathcal{D}_{tk}^+|\cdot\mu_{tk} \tag{14}$$

$$\mu_t = \sum_{k=1}^{K}\frac{|\mathcal{D}_{tk}^+|}{|\mathcal{D}_t^+|}\cdot\mu_{tk} \tag{15}$$

Now using the fact that $|\mathcal{D}_{tk}^+| = |\mathcal{D}_{tk}|p_{tk}$ and $|\mathcal{D}_t^+| = \sum_{k=1}^{K}|\mathcal{D}_{tk}|p_{tk}$:

$$\mu_t = \sum_{k=1}^{K}\frac{|\mathcal{D}_{tk}|p_{tk}}{\sum_{j=1}^{K}|\mathcal{D}_{tj}|p_{tj}}\cdot\mu_{tk} \tag{16}$$

$$\mu_t = \sum_{k=1}^{K} w_{tk}\cdot\mu_{tk} \tag{17}$$

Now that we've written μ_t in terms of $\mu_{t1}, \ldots, \mu_{tK}$, we'll switch our focus to showing that $\mathbb{E}_{x\sim\mathcal{D}_t^+}[\hat{\mu}_t]$ simplifies to this expression for μ_t. Plugging in the definition of our estimator, we have:

$$\mathbb{E}_{x\sim\mathcal{D}_t^+}[\hat{\mu}_t] = \mathbb{E}_{x\sim\mathcal{D}_t^+}\left[\sum_{k=1}^{K} w_{tk}\cdot\frac{\sum_{x\in X_{tk}^+} f(x)}{|X_{tk}^+|}\right] \tag{18}$$

Because the cardinality of $\mathcal{D}_{t1}^+, \ldots, \mathcal{D}_{tK}^+$ is independent of our sampling $X_t^+ \sim \mathcal{D}_t^+$ – and because we assume all p_{tk} are known – we can move w_{tk} outside of our expectation:

$$\mathbb{E}_{x\sim\mathcal{D}_t}[\hat{\mu}_t] = \sum_{k=1}^{K} w_{tk}\cdot\mathbb{E}_{x\sim X_{tk}^+}\left[\frac{\sum_{x\in X_{tk}^+} f(x)}{|X_{tk}^+|}\right] \tag{19}$$

Furthermore, because our allocations a_{tk} are assumed to be fixed and independent of our sampling – and because $|X_{tk}^+| = p_{tk}\left(\frac{N_1}{K} + N_2 a_{tk}\right)$ – this term can also be moved outside of the expectation:

$$\mathbb{E}_{x\sim\mathcal{D}_t}[\hat{\mu}_t] = \sum_{k=1}^{K} w_{tk}\cdot\frac{1}{|X_{tk}^+|}\cdot\sum_{x\in X_{tk}^+}\mathbb{E}_{x\sim\mathcal{S}_{tk}}[f(x)] \tag{20}$$

Each sample $x \in X_{tk}^+$ is drawn i.i.d. from $\mathcal{D}_{tk}^+$ under reservoir sampling. Therefore, the expectation $\mathbb{E}_{x\sim X_{tk}^+}[f(x)]$ is equal to the true mean μ_{tk} of the statistic function for the records in $\mathcal{D}_{tk}^+$ and we have:

$$\mathbb{E}_{x\sim\mathcal{D}_t^+}[\hat{\mu}_t] = \sum_{k=1}^{K} w_{tk}\cdot\frac{1}{|X_{tk}^+|}\cdot\sum_{x\in X_{tk}^+}\mu_{tk} \tag{21}$$

$$= \sum_{k=1}^{K} w_{tk}\cdot\frac{1}{|X_{tk}^+|}\cdot|X_{tk}^+|\cdot\mu_{tk} \tag{22}$$

$$= \sum_{k=1}^{K} w_{tk}\cdot\mu_{tk} \tag{23}$$

Looking back at equation (17), we can see that this is equal to μ_t. Therefore we have shown that $\mathbb{E}_{x\sim\mathcal{D}_t^+}[\hat{\mu}_t] = \mu_t$, which means that our estimator $\hat{\mu}_t$ is unbiased.

Lemma 2. The variance of our estimator $\hat{\mu}_t$ is $Var[\hat{\mu}_t] = \sum_{k=1}^{K} \frac{w_{tk}^2\sigma_{tk}^2}{p_{tk}\left(\frac{N_1}{K}+N_2a_{tk}\right)}$.

Proof.
Plugging in the definition of our estimator $\hat{\mu}_t$ we get:

$$Var[\hat{\mu}_t] = Var\left[\sum_{k=1}^{K} w_{tk} \cdot \frac{\sum_{x\in X_{tk}^+} f(x)}{|X_{tk}^+|}\right] \tag{24}$$

$$= \sum_{k=1}^{K} Var\left[w_{tk} \cdot \frac{\sum_{x\in X_{tk}^+} f(x)}{|X_{tk}^+|}\right] \tag{25}$$

By the property that $Var(aX) = a^2Var(X)$:

$$Var[\hat{\mu}_t] = \sum_{k=1}^{K} \left(\frac{w_{tk}}{|X_{tk}^+|}\right)^2 \cdot \sum_{x\in X_{tk}^+} Var[f(x)] \tag{26}$$

Each sample $x \in X_{tk}^+$ is drawn i.i.d. from $\mathcal{D}_{tk}^+$ under reservoir sampling. Therefore, the variance $Var[f(x)]$ is equal to the true variance σ_{tk}^2 of the statistic function for the records in $\mathcal{D}_{tk}^+$ and we have:

$$Var[\hat{\mu}_t] = \sum_{k=1}^{K} \left(\frac{w_{tk}}{|X_{tk}^+|}\right)^2 \cdot \sum_{x\in X_{tk}^+} \sigma_{tk}^2 \tag{27}$$

$$= \sum_{k=1}^{K} \left(\frac{w_{tk}}{|X_{tk}^+|}\right)^2 \cdot |X_{tk}^+| \cdot \sigma_{tk}^2 \tag{28}$$

$$= \sum_{k=1}^{K} \frac{w_{tk}^2\sigma_{tk}^2}{|X_{tk}^+|} \tag{29}$$

Finally, plugging in $|X_{tk}^+| = p_{tk}\left(\frac{N_1}{K} + N_2a_{tk}\right)$ gives us our expected error for $\hat{\mu}_t$:

$$Var[\hat{\mu}_t] = \sum_{k=1}^{K} \frac{w_{tk}^2\sigma_{tk}^2}{p_{tk}\left(\frac{N_1}{K} + N_2a_{tk}\right)} \tag{30}$$

From Lemma 1, we know that our estimator $\hat{\mu}_t$ is unbiased. Thus, our expected error will be equal to the variance term we proved in Lemma 2:

$$\mathbb{E}[(\hat{\mu}_t - \mu_t)^2] = Var[\hat{\mu}_t] \tag{31}$$

$$\mathbb{E}[(\hat{\mu}_t - \mu_t)^2] = \sum_{k=1}^{K} \frac{w_{tk}^2\sigma_{tk}^2}{p_{tk}\left(\frac{N_1}{K} + N_2a_{tk}\right)} \tag{32}$$

4 Analysis: Optimal Stratified Sampling Allocation

In this section, let us assume that we are given a query with a predicate. We wish to compute the optimal allocation of our dynamic sampling budget N_2. For this derivation we will assume that all σ_{tk} and p_{tk} are known, that $\mathbb{E}[\hat{\mu}_{tk}] = \mu_{tk}$, and that $|X_{tk}^+| = p_{tk}\left(\frac{N_1}{K} + N_2a_{tk}\right)$.

Theorem 2. Given the assumptions stated above, the optimal allocation a_{tk}^* of our dynamic sampling N_2 is:

$$a^*_{tk} = \frac{|\mathcal{D}_{tk}|\sqrt{p_{tk}}\sigma_{tk}}{\frac{N_2}{N}\sum_{j=1}^{K}|\mathcal{D}_{tj}|\sqrt{p_{tj}}\sigma_{tj}} - \frac{N_1}{N_2 K} \tag{33}$$

Proof. We will use the method of Lagrange multipliers to determine the choice of a_{tk} such that $a_{tk} > 0$ and $\sum_{k=1}^{K} a_{tk} = 1$ which minimizes the loss function $\mathcal{L}(a_{t1}, \ldots, a_{tK}) = \sum_{k=1}^{K} \frac{w_{tk}^2\sigma_{tk}^2}{p_{tk}\left(\frac{N_1}{K}+N_2 a_{tk}\right)}$ of the unbiased estimator $\hat{\mu}_t$.

We define our loss function $\mathcal{L}$ and equality constraint g as follows:

$$\mathcal{L}(a_{t1}, \ldots, a_{tK}) = \sum_{k=1}^{K} \frac{w_{tk}^2\sigma_{tk}^2}{p_{tk}\left(\frac{N_1}{K} + N_2 a_{tk}\right)} \tag{34}$$

$$g(a_{t1}, \ldots, a_{tK}) = \sum_{k=1}^{K} a_{tk} - 1 \tag{35}$$

Under the method of Lagrange multipliers, we need to solve the following system of equations:

$$\frac{\partial \mathcal{L}}{\partial a_{t1}} = \lambda \frac{\partial g}{\partial a_{t1}} \tag{36}$$

$$\ldots \tag{37}$$

$$\frac{\partial \mathcal{L}}{\partial a_{tK}} = \lambda \frac{\partial g}{\partial a_{tK}} \tag{38}$$

$$g(a_{t1}, \ldots, a_{tK}) = 1 \tag{39}$$

Let us first consider the gradient equation for some individual allocation a_{ti}. Plugging $\mathcal{L}$ and g into their respective partial derivatives gives us:

$$\frac{\partial}{\partial a_{ti}}\left(\sum_{k=1}^{K} \frac{w_{tk}^2\sigma_{tk}^2}{p_{tk}\left(\frac{N_1}{K} + N_2 a_{tk}\right)}\right) = \lambda \frac{\partial}{\partial a_{ti}}\left(\sum_{k=1}^{K} a_{tk} - 1\right) \tag{40}$$

All terms in the summations that do not contain a_{ti} will drop out, thus our equation simplifies to:

$$-\frac{w_{ti}^2\sigma_{ti}^2 N_2}{p_{ti}\left(\frac{N_1}{K} + N_2 a_{ti}\right)^2} = \lambda \tag{41}$$

Because λ is a constant we can redefine $\lambda = -\lambda$ to get:

$$\frac{w_{ti}^2\sigma_{ti}^2 N_2}{p_{ti}\left(\frac{N_1}{K} + N_2 a_{ti}\right)^2} = \lambda \tag{42}$$

Furthermore, because of the symmetry of our gradient equations we will get the same result $\forall i \in 1, \ldots, K$. Solving for a_{ti} gives us:

$$a_{ti} = \frac{w_{ti}\sigma_{ti}}{\sqrt{\lambda N_2 p_{ti}}} - \frac{N_1}{N_2 K} \tag{43}$$

Turning our focus to the constraint equation and plugging in our result above we get:

$$\sum_{k=1}^{K} a_{tk} = 1 \tag{44}$$

$$\sum_{k=1}^{K}\left(\frac{w_{tk}\sigma_{tk}}{\sqrt{\lambda N_2 p_{ti}}} - \frac{N_1}{N_2 K}\right) = 1 \tag{45}$$

Solving for λ gives us:

$$\lambda = \left(\sum_{k=1}^{K} \frac{w_{tk}\sigma_{tk}\sqrt{N_2}}{\sqrt{p_{tk}}N} \right)^2 \tag{46}$$

Now that we've solved for the Lagrange multiplier, we can plug λ back into our equation for a_{tk} to get the optimal a^*_{tk}:

$$a^*_{tk} = \frac{w_{tk}\sigma_{tk}}{\sqrt{N_2 p_{tk}}\left(\sum_{j=1}^{K} \frac{w_{tj}\sigma_{tj}\sqrt{N_2}}{\sqrt{p_{tj}}N}\right)} - \frac{N_1}{N_2 K} \tag{47}$$

$$a^*_{tk} = \frac{w_{tk}\sigma_{tk}}{\frac{N_2}{N}\sqrt{p_{tk}}\sum_{j=1}^{K} \frac{w_{tj}\sigma_{tj}}{\sqrt{p_{tj}}}} - \frac{N_1}{N_2 K} \tag{48}$$

The w_{tk} and w_{tj} in our expression share a constant denominator which can be cancelled out, leaving us with:

$$a^*_{tk} = \frac{|\mathcal{D}_{tk}|p_{tk}\sigma_{tk}}{\frac{N_2}{N}\sqrt{p_{tk}}\sum_{j=1}^{K} \frac{\mathcal{D}_{tj}|p_{tj}\sigma_{tj}}{\sqrt{p_{tj}}}} - \frac{N_1}{N_2 K} \tag{49}$$

Simplifying one step further, we can rewrite the equation above as:

$$a^*_{tk} = \frac{|\mathcal{D}_{tk}|\sqrt{p_{tk}}\sigma_{tk}}{\frac{N_2}{N}\sum_{j=1}^{K} |\mathcal{D}_{tj}|\sqrt{p_{tj}}\sigma_{tj}} - \frac{N_1}{N_2 K} \tag{50}$$

Our loss function $\mathcal{L}$ is unbounded above, as it could grow arbitrarily large with $a_{tk} \to 0$. Our solution for λ yielded a single value, thus we can conclude that since $\mathcal{L}$ has no maximum this λ must correspond to a minimum.

5 Analysis: Convergence to Optimal Allocation

In this section, let us assume that we are given a query with a predicate and let us assume that our dataset follows a stationary distribution. Specifically, we assume that:

$$\sigma_{tk} = \sigma_{rk} \ \forall t, r \in [1, T]$$
$$w_{tk} = w_{rk} \ \forall t, r \in [1, T]$$
$$p_{tk} = p_{rk} \ \forall t, r \in [1, T]$$

In the pilot segment ($t = 1$) we perform uniform sampling. For segments $t \geq 2$ we define the sample allocations $\hat{a}_{tk}$ to be computed as follows:

$$\hat{a}_{tk} = \frac{|\mathcal{D}_{<tk}|\sqrt{\hat{p}_{<tk}}\hat{\sigma}_{<tk}}{\frac{N_2}{N}\sum_{j=1}^{K} |\mathcal{D}_{<tj}|\sqrt{\hat{p}_{<tj}}\hat{\sigma}_{<tj}} - \frac{N_1}{N_2 K} \tag{51}$$

Theorem 3. Given the assumptions stated above:

$$\mathbb{E}[(\hat{a}_{tk} - a^*_{tk})^2] \leq O\left(\frac{1}{N_1(t-1)}\right) \tag{52}$$

Proof. From equation (51), we can see that $\hat{a}_{tk}$ is a function of two random variables: $\hat{p}_{<tk}$ and $\hat{\sigma}_{<tk}$. Thus, we'll begin our proof by putting concentration inequalities on $\hat{p}_{<tk}$ and $\hat{\sigma}_{<tk}$ in four separate lemmas. We'll then apply these concentration inequalities in our expression for $\mathbb{E}[(\hat{a}_{tk} - a^*_{tk})^2]$ to derive the rate of convergence.

5.0.1 Concentration Inequality on $\hat{p}_{<tk}$

Lemma 5.1 ($\sqrt{\hat{p}_{<tk}}$ Upper Bound). *For small $\delta > 0$, the following holds for all k simultaneously,*

$$\sqrt{\hat{p}_{<tk}} \leq \sqrt{p_{<tk}} \cdot \sqrt{1 + O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)} \tag{53}$$

Proof. Let us first define $N_{<t} = (t-1)N_1/K$. We apply Chung and Lu [1] reproduced as Lemma 7.1 to upper bound $\hat{p}_{<tk}$. We set $Y_i \sim Bernoulli(p_{<tk})$ and $a_i = \frac{1}{N_{<t}}$ for all $i \in [N_{<t}]$. Thus $v = \sum_{i=1}^{N_{<t}} a_i^2 p_i = \frac{p_{<tk}}{N_{<t}}$ and $Y = \sum_{i=1}^{N_{<t}} a_i Y_i = \hat{p}_{<tk}$. Thus, according to the Lemma,

$$P(\hat{p}_{tk} \geq p_{tk} + \lambda) \leq exp\left(\frac{-\lambda^2}{2(\frac{p_{<tk}}{N_{<t}} + \frac{\lambda}{3N_{<t}})}\right) \tag{54}$$

We will set $\delta = exp\big(-\lambda^2/2(\frac{p_{<tk}}{N_{<t}} + \frac{\lambda}{3N_{<t}})\big)$ and $\lambda = \big(\sqrt{18N_{<t}p_{<tk}\ln(1/\delta) + \ln(1/\delta)^2} - \ln(1/\delta)\big)/\big(3N_{<t}\big)$. Thus, with probability at least $1-\delta$,

$$\begin{aligned}\hat{p}_{<tk} &\leq p_{<tk} + \frac{\sqrt{18N_{<t}p_{<tk}\ln(1/\delta) + \ln(1/\delta)^2} - \ln(1/\delta)}{3N_{<t}} &(55)\\ &\leq p_{<tk} + \frac{\sqrt{18N_{<t}p_{<tk}\ln(1/\delta)}}{3N_{<t}} = p_{<tk} + \sqrt{\frac{2\ln(1/\delta)p_{<tk}}{N_{<t}}} &(56)\end{aligned}$$

With probability at least $1-\delta$,

$$\begin{aligned}\sqrt{\hat{p}_{<tk}} &\leq \sqrt{p_{<tk} + \sqrt{\frac{2\ln(1/\delta)p_{<tk}}{N_{<t}}}} \leq \sqrt{p_{<tk}} \cdot \sqrt{1 + \sqrt{\frac{2\ln(1/\delta)}{p_{<tk}N_{<t}}}} &(57)\\ \sqrt{\hat{p}_{<tk}} &\leq \sqrt{p_{<tk}} \cdot \sqrt{1 + O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)} &(58)\end{aligned}$$

We union bound across all strata $S_{t1}, \ldots, S_{tk}$. As a result, the bound above holds for all k simultaneously with probability at least $1 - K\delta$.

Lemma 5.2 ($\sqrt{\hat{p}_{<tk}}$ Lower Bound). *For small $\delta > 0$, the following holds for all k simultaneously,*

$$\sqrt{\hat{p}_{<tk}} \geq \sqrt{p_{<tk}} \cdot \sqrt{1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)} \tag{59}$$

Proof. Let us first define $N_{<t} = (t-1)N_1/K$. We apply Chung and Lu [1] reproduced as Lemma 7.1 to lower bound $\hat{p}_{<tk}$. We set $Y_i \sim Bernoulli(p_{<tk})$ and $a_i = \frac{1}{N_{<t}}$ for all $i \in [N_{<t}]$. Thus $v = \sum_{i=1}^{N_{<t}} a_i^2 p_i = \frac{p_{<tk}}{N_{<t}}$ and $Y = \sum_{i=1}^{N_{<t}} a_i Y_i = \hat{p}_{<tk}$. Thus, according to the Lemma,

$$P(\hat{p}_{tk} < p_{tk} - \lambda) \leq exp\left(\frac{-\lambda^2}{2(\frac{p_{<tk}}{N_{<t}})}\right) \tag{60}$$

We will set $\delta = exp\left(\frac{-\lambda^2}{2(\frac{p_{<tk}}{N_{<t}})}\right)$ and $\lambda = \frac{\sqrt{p_{<tk}}}{\sqrt{N_{<t}}} \cdot \sqrt{2\ln(1/\delta)}$. Thus, with probability at least $1-\delta$,

$$\hat{p}_{<tk} \geq p_{<tk} - \frac{\sqrt{p_{<tk}}}{\sqrt{N_{<t}}} \cdot \sqrt{2\ln(1/\delta)} = p_{<tk} - \sqrt{\frac{2\ln(1/\delta)p_{<tk}}{N_{<t}}} \tag{61}$$

In other words, with probability at least $1 - \delta$,

$$\sqrt{\hat{p}_{<tk}} \geq \sqrt{p_{<tk} - \sqrt{\frac{2\ln(1/\delta)p_{<tk}}{N_{<t}}}} \geq \sqrt{p_{<tk}}\sqrt{1 - \sqrt{\frac{2\ln(1/\delta)}{p_{<tk}N_{<t}}}} \tag{62}$$

$$\geq \sqrt{p_{<tk}}\sqrt{1 - O\Bigg(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\Bigg)} \tag{63}$$

We then apply the union bound, so the above holds for all k simultaneously with probability at least $1 - K\delta$.

5.0.2 Concentration Inequality on $\hat{\sigma}_{<tk}$

Before we can prove our concentration inequality on $\hat{\sigma}_{<tk}$, we first need to derive a lower bound on the number of predicate matching samples $|X^+_{<tk}|$.

Lemma 5.3 ($|X^+_{<tk}|$ Lower Bound). *For small $\delta > 0$, the following is true for all k,*

$$|X^+_{<tk}| \geq p_{<tk}N_{<t} - \sqrt{2\ln(1/\delta)p_{<tk}N_{<t}} \tag{64}$$

Proof. Let us first define $N_{<t} = (t-1)N_1/K$. We apply Tarjan [4] reproduced as Lemma 7.2 to lower bound $|X^+_{<tk}|$. Note that $|X^+_{<tk}| \sim Binomial(N_{<t}, p_{<tk})$ and that equivalently $|X^+_{<tk}| = \sum_{i=1}^{N_{<t}} Y_i$ where $Y_i \sim Bernoulli(p_{<tk})$. Hence,

$$P\Big(|X^+_{<tk}| \leq (1-\epsilon)p_{<tk}N_{<t}\Big) \leq exp\Big(\frac{-\epsilon^2 p_{<tk}N_{<t}}{2}\Big) \tag{65}$$

We set $\delta = exp\Big(\frac{-\epsilon^2 p_{<tk}N_{<t}}{2}\Big)$ and $\epsilon = \sqrt{\frac{2\ln(1/\delta)}{p_{<tk}N_{<t}}}$. Thus, with probability at least $1 - \delta$:

$$|X^+_{<tk}| \geq p_{<tk}N_{<t} - \sqrt{2\ln(1/\delta)p_{<tk}N_{<t}} \tag{66}$$

We union bound across all k. As a result, with probability at least $1 - K\delta$ the above holds for all k simultaneously.

Lemma 5.4 ($\hat{\sigma}_{<tk}$ Upper and Lower Bound for $|X^+_{<tk}| \geq 2$). *For small $\delta > 0$, the following holds for all k simultaneously,*

$$\sigma_{<tk}\sqrt{1 - O\Bigg(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\Bigg)} \leq \hat{\sigma}_{<tk} \leq \sigma_{<tk}\sqrt{1 + O\Bigg(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\Bigg)} \tag{67}$$

Proof. Let us first define $N_{<t} = (t-1)N_1/K$. To obtain a concentration inequality on $\sigma_{<tk}$, we will apply bounded differences on the unbiased sample variance estimator for $|X^+_{<tk}| \geq 2$, which is:

$$\hat{\sigma}^2_{tk} = \frac{1}{|X^+_{<tk}| - 1}\sum_{i=1}^{|X^+_{<tk}|}(X_{tk,i} - \hat{\mu}_{tk})^2 \tag{68}$$

Note that the unbiased sample variance is a U-statistic U_n that arises from taking $g(X_{tk,i}, X_{tk,j}) = \frac{1}{2}(X_{tk,i} - X_{tk,j})^2$ (Ferguson [2] reproduced as Lemma 7.3). For bounded U-statistics, we can apply bounded differences to obtain a concentration bound (Rinaldo [3] reproduced as Lemma 7.4). Namely, where $g(X_{tk,i}, X_{tk,j}) \leq b$,

$$P(|U_n - \mathbb{E}[U_n]| \geq t) \leq 2\ \exp\Big(-\frac{nt^2}{8b^2}\Big) \tag{69}$$

We use this concentration inequality with $n = |X^+_{<tk}|$ and $g(X_{tk,i}, X_{tk,j}) \leq b = C_k^{(\mu_t^2)}$. Hence,

$$P(|\hat{\sigma}_{<tk} - \sigma_{<tk}| \geq t) \leq 2\, \exp\bigg(- \frac{|X^+_{<tk}|t^2}{8C_k^{(\mu_t^4)}} \bigg) \tag{70}$$

We then set $2\delta = 2\, \exp\Big(- \frac{|X^+_{<tk}|t^2}{8C_k^{(\mu_t^4)}} \Big)$ and $t = \sqrt{8\ln(1/\delta) C_k^{(\mu_t^4)} / |X^+_{<tk}|}$. Thus, with probability at least $1 - 2\delta$:

$$\sigma^2_{<tk} - \sqrt{\frac{8\ln(1/\delta) C_k^{(\mu_t^4)}}{|X^+_{<tk}|}} \leq \hat{\sigma}^2_{<tk} \leq \sigma^2_{<tk} + \sqrt{\frac{8\ln(1/\delta) C_k^{(\mu_t^4)}}{|X^+_{<tk}|}} \tag{71}$$

Using Lemma 5.3, we have $|X^+_{<tk}| \geq p_{<tk}N_{<t} - \sqrt{2\ln(1/\delta) p_{<tk} N_{<t}}$ with probability at least $1 - K\delta$. Note that if that bound holds, then $p_{<tk} \geq p^*_{<} = \frac{2\ln(1/\delta) + 2\sqrt{\ln(1/\delta)} + 2}{|X^+_{<tk}|}$, which allows us to simplify (75) into (76):

$$\sigma^2_{<tk} - \sqrt{\frac{8\ln(1/\delta) C_k^{(\mu_t^4)}}{p_{<tk}N_{<t} - \sqrt{2\ln(1/\delta) p_{<tk} N_{<t}}}} \leq \hat{\sigma}^2_{<tk} \leq \sigma^2_{<tk} + \sqrt{\frac{8\ln(1/\delta) C_k^{(\mu_t^4)}}{p_{<tk}N_{<t} - \sqrt{2\ln(1/\delta) p_{<tk} N_{<t}}}} \tag{72}$$

$$\sigma^2_{<tk} - \sqrt{\frac{8\ln(1/\delta) C_k^{(\mu_t^4)}}{p_{<tk}N_{<t}\big(1 - \sqrt{\frac{2\ln(1/\delta)}{p_{<tk}N_{<t}}}\big)}} \leq \hat{\sigma}^2_{<tk} \leq \sigma^2_{<tk} + \sqrt{\frac{8\ln(1/\delta) C_k^{(\mu_t^4)}}{p_{<tk}N_{<t}\big(1 - \sqrt{\frac{2\ln(1/\delta)}{p_{<tk}N_{<t}}}\big)}} \tag{73}$$

$$\sigma^2_{<tk} - \sqrt{\frac{8\ln(1/\delta) C_k^{(\mu_t^4)}}{p_{<tk}N_{<t}} \cdot \bigg(1 + \sqrt{\frac{2\ln(1/\delta)}{p_{<tk}N_{<t}}}\bigg)} \leq \hat{\sigma}^2_{<tk} \leq \sigma^2_{<tk} + \sqrt{\frac{8\ln(1/\delta) C_k^{(\mu_t^4)}}{p_{<tk}N_{<t}} \cdot \bigg(1 + \sqrt{\frac{2\ln(1/\delta)}{p_{<tk}N_{<t}}}\bigg)} \tag{74}$$

$$\sigma^2_{<tk} - \sqrt{\frac{8\ln(1/\delta) C_k^{(\mu_t^4)}}{p_{<tk}N_{<t}} + \frac{8\sqrt{2}\ln(1/\delta)^{3/2}}{p^{3/2}_{<tk}N^{3/2}_{<t}}} \leq \hat{\sigma}^2_{<tk} \leq \sigma^2_{<tk} + \sqrt{\frac{8\ln(1/\delta) C_k^{(\mu_t^4)}}{p_{<tk}N_{<t}} + \frac{8\sqrt{2}\ln(1/\delta)^{3/2}}{p^{3/2}_{<tk}N^{3/2}_{<t}}} \tag{75}$$

$$\sigma^2_{<tk} - \sqrt{O\bigg(\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}\bigg)} \leq \hat{\sigma}^2_{<tk} \leq \sigma^2_{<tk} + \sqrt{O\bigg(\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}\bigg)} \tag{76}$$

Finally, we can take the square root of the inequality to get our expression in terms of $\sigma_{<tk}$:

$$\sqrt{\sigma^2_{<tk} - \sqrt{O\bigg(\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}\bigg)}} \leq \hat{\sigma}_{<tk} \leq \sqrt{\sigma^2_{<tk} + \sqrt{O\bigg(\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}\bigg)}} \tag{77}$$

$$\sigma_{<tk}\sqrt{1 - O\bigg(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\bigg)} \leq \hat{\sigma}_{<tk} \leq \sigma_{<tk}\sqrt{1 + O\bigg(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\bigg)} \tag{78}$$

We union bound across the strata and intersection bound our application of Lemma 5.3, and get that the above holds with probability at least $(1 - 2K\delta)(1 - K\delta) = 1 - 3K\delta + 2K^2\delta^2 \geq 1 - 3K\delta$. The derivation above relies on $\sigma_{<tk} \neq 0$ because we divide by $\sigma_{<tk}$ in (77). However, if $\sigma_{<tk} = 0$, then $\hat{\sigma}_{<tk} = \sigma_{<tk}$, so the bound still holds.

5.0.3 Final Result

We'll now use the concentration inequalities on $\hat{p}_{<tk}$ and $\hat{\sigma}_{<tk}$ to derive an upper bound for $\mathbb{E}[(\hat{a}_{tk} - a^*_{tk})^2]$. Plugging in for $\hat{a}_{tk}$ and a^*_{tk} and doing some minor simplifying we have:

$$\mathbb{E}[(\hat{a}_{tk} - a^*_{tk})^2] = \mathbb{E}\Bigg[\bigg(\frac{|\mathcal{D}_{<tk}|N}{N_2}\bigg(\frac{\sqrt{\hat{p}_{<tk}}\hat{\sigma}_{<tk}}{\sum_j |\mathcal{D}_{<tj}|\sqrt{\hat{p}_{<tj}}\hat{\sigma}_{<tj}} - \frac{\sqrt{p_{<tk}}\sigma_{<tk}}{\sum_j |\mathcal{D}_{<tj}|\sqrt{p_{<tj}}\sigma_{<tj}}\bigg)\bigg)^2\Bigg] \tag{79}$$

We are going to derive the upper bound for our expression by upper bounding the numerator and lower bounding the denominator in the term with $\hat{p}_{<tk}$ and $\hat{\sigma}_{<tk}$. We first upper bound the numerator. We split our analysis into the cases where $|X^+_{<tk}|$ is large and small. For large $|X^+_{<tk}|$ we have:

$$\sqrt{\hat{p}_{<tk}}\hat{\sigma}_{<tk} \leq \sqrt{p_{<tk}}\sigma_{<tk}\sqrt{1+O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)}\sqrt{1+O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)} \tag{80}$$

$$\leq \sqrt{p_{<tk}}\sigma_{<tk}\left(1+O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)\right) \tag{81}$$

$$\leq \sqrt{p_{<tk}}\sigma_{<tk}+O\left(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\right) \tag{82}$$

In Lemma 6.8, we derive a similar expression for cases where $|X^+_{<tk}| < 2$. Next, we lower bound the denominator:

$$\sum_{k=1}^{K}|\mathcal{D}_{<tk}|\sqrt{\hat{p}_{<tk}}\hat{\sigma}_{<tk} \geq \sum_{p_{<tk}\geq p^*_<}|\mathcal{D}_{<tk}|\sqrt{p_{<tk}}\sigma_{<tk}\left(1-O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)\right)+\sum_{p_{<tk}<p^*_<}|\mathcal{D}_{<tk}|\sqrt{\hat{p}_{<tk}}\hat{\sigma}_{<tk}O\left(\frac{1}{\sqrt{N_{<t}}}\right)\Bigg) \tag{83}$$

$$\geq \sum_{k=1}^{K}|\mathcal{D}_{<tk}|\sqrt{p_{<tk}}\sigma_{<tk}-O\left(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\right) \tag{84}$$

Plugging these back into our upper bound for $\mathbb{E}[(\hat{a}_{tk}-a^*_{tk})^2]$ we have:

$$\mathbb{E}[(\hat{a}_{tk}-a^*_{tk})^2] \leq \left(\frac{|\mathcal{D}_{<tk}|N}{N_2}\left(\frac{\sqrt{p_{<tk}}\sigma_{<tk}+O\left(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\right)}{\sum_j|\mathcal{D}_{<tj}|\sqrt{p_{<tj}}\sigma_{<tj}-O\left(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\right)}-\frac{\sqrt{p_{<tk}}\sigma_{<tk}}{\sum_j|\mathcal{D}_{<tj}|\sqrt{p_{<tj}}\sigma_{<tj}}\right)\right)^2 \tag{85}$$

Multiplying the numerator and denominator of the term with the Big-O notation by $\frac{1}{\sum_k|\mathcal{D}_{<tk}|p_{<tk}\sigma_{<tk}}$, substituting using our expression for a^*_{tk} in (50), and simplifying gives us:

$$\leq \left(\frac{|\mathcal{D}_{<tk}|N}{N_2}\left(\frac{(a^*_{tk}+\frac{N_1}{N_2K})(\frac{N_2}{N|\mathcal{D}_{<tk}|})+O\left(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\right)}{1-O\left(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\right)}-\left(a^*_{tk}+\frac{N_1}{N_2K}\right)\left(\frac{N_2}{N|\mathcal{D}_{<tk}|}\right)\right)\right)^2 \tag{86}$$

$$\leq \left(\frac{|\mathcal{D}_{<tk}|N}{N_2}\left(\left((a^*_{tk}+\frac{N_1}{N_2K})(\frac{N_2}{N|\mathcal{D}_{<tk}|})+O\left(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\right)\right)\left(1+O\left(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\right)\right)-\left(a^*_{tk}+\frac{N_1}{N_2K}\right)\left(\frac{N_2}{N|\mathcal{D}_{<tk}|}\right)\right)\right)^2 \tag{87}$$

$$\leq \left(\frac{|\mathcal{D}_{<tk}|N}{N_2}\left(\left(a^*_{tk}+\frac{N_1}{N_2K}\right)\left(\frac{N_2}{N|\mathcal{D}_{<tk}|}\right)+O\left(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\right)-\left(a^*_{tk}+\frac{N_1}{N_2K}\right)\left(\frac{N_2}{N|\mathcal{D}_{<tk}|}\right)\right)\right)^2 \tag{88}$$

$$\leq \left(\left(a^*_{tk}+\frac{N_1}{N_2K}\right)+O\left(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\right)-\left(a^*_{tk}+\frac{N_1}{N_2K}\right)\right)^2 \tag{89}$$

$$\leq \left(O\left(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\right)\right)^2 \tag{90}$$

$$\leq O\left(\frac{1}{N_{<t}}\right) \tag{91}$$

Finally, substituting in the equation for $N_{<t}$ gives us:

$$\mathbb{E}[(\hat{a}_{tk} - a^*_{tk})^2] \leq O\left(\frac{1}{N_1(t-1)}\right) \tag{92}$$

6 Upper Bound on Expected Error

We wish to compute an upper bound for our expected error. In other words, we wish to compute an upper bound on the expression:

$$\mathbb{E}[(\hat{\mu}_t - \mu_t)^2] \tag{93}$$

We assume that we are given a query with a predicate and let us assume that our dataset follows a stationary distribution. Specifically, we assume that:

$$\sigma_{tk} = \sigma_{rk} \ \forall t, r \in [1, T]$$
$$w_{tk} = w_{rk} \ \forall t, r \in [1, T]$$
$$p_{tk} = p_{rk} \ \forall t, r \in [1, T]$$

In the pilot segment ($t = 1$) we perform uniform sampling. For segments $t \geq 2$ we define the sample allocations $\hat{a}_{tk}$ to be computed as follows:

$$\hat{a}_{tk} = \frac{|\mathcal{D}_{<tk}|\sqrt{\hat{p}_{<tk}}\hat{\sigma}_{<tk}}{\frac{N_2}{N}\sum_{j=1}^{K}|\mathcal{D}_{<tj}|\sqrt{\hat{p}_{<tj}}\hat{\sigma}_{<tj}} - \frac{N_1}{N_2 K} \tag{94}$$

Theorem 4. Given the assumptions stated above, the upper bound on the expected error of our estimator $\hat{\mu}_t$ is:

$$\mathbb{E}[(\hat{\mu}_t - \mu_t)^2] \leq O\left(\frac{1}{N_1}\right) + O\left(\frac{N_1}{N_2^2}\right) + O\left(\frac{1}{N_2\sqrt{N_1}}\right) + O\left(\frac{1}{N_2\sqrt{N_1}t}\right) \tag{95}$$

Proof. We'll begin by proving a set of lemmas that will help us in our final proof. We'll first place high probability bounds on all of our random variables. We'll then place upper or lower bounds as appropriate on other quantities of interest.

6.1 Additional Lemmas

Lemma 6.5 ($\sqrt{\hat{p}_{tk}}$ Upper Bound). *For small $\delta > 0$, the following holds for all k simultaneously,*

$$\sqrt{\hat{p}_{tk}} \leq \sqrt{p_{tk}} \cdot \sqrt{1 + O\left(\sqrt{\frac{\ln(1/\delta)}{p_{tk}N_1}}\right)} \tag{96}$$

Proof. This proof is identical to the one used in Lemma 5.1, except instead of setting $n = N_{<t}$ we set $n = N_1/K$.

Lemma 6.6 ($\sqrt{\hat{p}_{tk}}$ Lower Bound). *For small $\delta > 0$, the following holds for all k simultaneously,*

$$\sqrt{\hat{p}_{tk}} \geq \sqrt{p_{tk}} \cdot \sqrt{1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{tk}N_1}}\right)} \tag{97}$$

Proof. This proof is identical to the one used in Lemma 5.2, except instead of setting $n = N_{<t}$ we set $n = N_1/K$.

Lemma 6.7 ($\hat{w}_{tk}$ Upper and Lower Bounds). *For small $\delta > 0$, the bounds of Lemma 6.5 and 6.6 hold and,*

$$w_{tk} - O\Bigg(\frac{\ln(1/\delta)\sqrt{p_{tk}}}{\sqrt{N_1}}\Bigg) \leq \hat{w}_{tk} \leq w_{tk} + O\Bigg(\frac{\ln(1/\delta)\sqrt{p_{tk}}}{\sqrt{N_1}}\Bigg) \tag{98}$$

Proof. Recall that $\hat{w}_{tk} = |\mathcal{D}_{tk}|p_{tk}/\sum_{j=1}^{K}|\mathcal{D}_{tj}|p_{tj}$. Equation (98) holds on the event that the bounds of Lemma 6.5 and 6.6 hold for all k simultaneously, which occurs with probability at least $1 - 2K\delta$ according to the union bound. For the remainder of the proof of this Lemma, we condition on this event and prove (98). Specifically, the condition is that for each k,

$$p_{tk} - \sqrt{\frac{2\ln(1/\delta)p_{tk}}{N_1}} \leq \hat{p}_{tk} \leq p_{tk} + \sqrt{\frac{2\ln(1/\delta)p_{tk}}{N_1}} \tag{99}$$

We will now maximize or minimize the numerator and denominator appropriately in $\hat{w}_{tk}$:

$$\frac{|\mathcal{D}_{tk}|\big(p_{tk} - \sqrt{\frac{2\ln(1/\delta)p_{tk}}{N_1}}\big)}{\sum_{j=1}^{K}|\mathcal{D}_{tj}|\big(p_{tj} + \sqrt{\frac{2\ln(1/\delta)p_{tj}}{N_1}}\big)} \leq \hat{w}_{tk} \leq \frac{|\mathcal{D}_{tk}|\big(p_{tk} + \sqrt{\frac{2\ln(1/\delta)p_{tk}}{N_1}}\big)}{\sum_{j=1}^{K}|\mathcal{D}_{tj}|\big(p_{tj} - \sqrt{\frac{2\ln(1/\delta)p_{tj}}{N_1}}\big)} \tag{100}$$

We now simplify this bound with Big-O notation.

$$\begin{aligned}
\hat{w}_{tk} &\leq \frac{|\mathcal{D}_{tk}|\big(p_{tk} + \sqrt{\frac{2\ln(1/\delta)p_{tk}}{N_1}}\big)}{\sum_{j=1}^{K}|\mathcal{D}_{tj}|\big(p_{tj} - \sqrt{\frac{2\ln(1/\delta)p_{tj}}{N_1}}\big)} &(101)\\
&\leq \frac{|\mathcal{D}_{tk}|p_{tk}\big(1 + \sqrt{\frac{2\ln(1/\delta)}{p_{tk}N_1}}\big)}{\sum_{j=1}^{K}|\mathcal{D}_{tj}|p_{tj}\big(1 - \sqrt{\frac{2\ln(1/\delta)}{p_{tj}N_1}}\big)} &(102)\\
&\leq \frac{\frac{1}{\sum_{j=1}^{K}|\mathcal{D}_{tj}|p_{tj}} \cdot |\mathcal{D}_{tk}|p_{tk}\big(1 + \sqrt{\frac{2\ln(1/\delta)}{p_{tk}N_1}}\big)}{\frac{1}{\sum_{j=1}^{K}|\mathcal{D}_{tj}|p_{tj}} \cdot \sum_{j=1}^{K}|\mathcal{D}_{tj}|p_{tj}\big(1 - \sqrt{\frac{2\ln(1/\delta)}{p_{tj}N_1}}\big)} &(103)\\
&\leq \frac{w_{tk}\left(1 + \sqrt{\frac{2\ln(1/\delta)}{p_{tk}N_1}}\right)}{1 - \sqrt{\frac{2\ln(1/\delta)}{p_{tk}N_1}}} &(104)\\
&\leq w_{tk}\left(1 + \sqrt{\frac{2\ln(1/\delta)}{p_{tk}N_1}}\right)^2 &(105)\\
&\leq w_{tk}\left(1 + 2\sqrt{\frac{2\ln(1/\delta)}{p_{tk}N_1}} + \frac{2\ln(1/\delta)}{p_{tk}N_1}\right) &(106)\\
&\leq w_{tk} + O\Bigg(\sqrt{\frac{\ln(1/\delta)p_{tk}}{N_1}}\Bigg) + O\Bigg(\frac{\ln(1/\delta)}{N_1}\Bigg) &(107)\\
&\leq w_{tk} + O\Bigg(\frac{\ln(1/\delta)\sqrt{p_{tk}}}{\sqrt{N_1}}\Bigg) &(108)
\end{aligned}$$

We repeat these steps for the lower bound. Thus:

$$w_{tk} - O\Bigg(\frac{\ln(1/\delta)\sqrt{p_{tk}}}{\sqrt{N_1}}\Bigg) \leq \hat{w}_{tk} \leq w_{tk} + O\Bigg(\frac{\ln(1/\delta)\sqrt{p_{tk}}}{\sqrt{N_1}}\Bigg) \tag{109}$$

Lemma 6.8 ($\sqrt{\hat{p}_{<tk}}\hat{\sigma}_{<tk}$ Upper Bound for $|X^+_{<tk}| < 2$). *For $|X^+_{<tk}| < 2$, the following holds for all k:*

$$\sqrt{\hat{p}_{<tk}}\hat{\sigma}_{<tk} \leq \sqrt{p_{<tk}}\sigma_{<tk} + O\left(\frac{1}{\sqrt{N_{<t}}}\right) \tag{110}$$

Proof. In the cases where $|X^+_{<tk}| < 2$, we will upper bound $\sqrt{\hat{p}_{<tk}}\hat{\sigma}_{<tk}$ directly instead of combining our concentration inequalities for $\sqrt{\hat{p}_{<tk}}$ and $\hat{\sigma}_{<tk}$ separately.

$$\sqrt{\hat{p}_{<tk}}\hat{\sigma}_{<tk} \leq \sqrt{\frac{2}{N_{<t}}}\hat{\sigma}_{<tk} \leq \sqrt{p_{<tk}}\sigma_{<tk} + \sqrt{\frac{2}{N_{<t}}}C_k^{(\sigma)} \tag{111}$$

$$\sqrt{\hat{p}_{<tk}}\hat{\sigma}_{<tk} \leq \sqrt{p_{<tk}}\sigma_{<tk} + O\left(\frac{1}{\sqrt{N_{<t}}}\right) \tag{112}$$

Lemma 6.9 ($\hat{a}_{tk}$ Lower Bound). *For small $\delta > 0$, the following holds for all k where $p_{<tk} > p^*_< = \frac{2\ln(1/\delta)+2\sqrt{\ln(1/\delta)}+2}{|X^+_{<tk}|}$,*

$$\hat{a}_{tk} \geq a_{tk}\left(1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{tk}(t-1)N_1}}\right)\right) \tag{113}$$

Proof. Our goal is to lower bound:

$$\hat{a}_{tk} = \frac{|\mathcal{D}_{<tk}|\sqrt{\hat{p}_{<tk}}\hat{\sigma}_{<tk}}{\frac{N_2}{N}\sum_{j=1}^{K}|\mathcal{D}_{<tj}|\sqrt{\hat{p}_{<tj}}\hat{\sigma}_{<tj}} - \frac{N_1}{N_2 K} \tag{114}$$

To accomplish this, we will lower bound the non-constant term by upper bounding the denominator and lower bounding the numerator. To lower bound the numerator we will apply our concentration inequalities for $\sqrt{\hat{p}_{<tk}}$ and $\hat{\sigma}_{<tk}$ from Lemmas 5.2 and 5.4. We only lower bound for k where $|X^+_{<tk}| \geq 2$. Through Lemma 5.3, we know that for $|X^+_{<tk}| > 2$ to be true, we can condition on $p_{<tk} > p^*_< = \frac{2\ln(1/\delta)+2\sqrt{\ln(1/\delta)}+2}{|X^+_{<tk}|}$. This allows us to satisfy the conditions of Lemma 5.3.

$$\sqrt{\hat{p}_{<tk}}\hat{\sigma}_{<tk} \geq \sqrt{p_{<tk}}\sigma_{<tk}\sqrt{1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)}\sqrt{1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)} \tag{115}$$

$$\geq \sqrt{p_{<tk}}\sigma_{<tk}\left(1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)\right) \tag{116}$$

$$\geq \sqrt{p_{<tk}}\sigma_{<tk} - O\left(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\right) \tag{117}$$

To upper bound the denominator, we also apply Lemma 5.1 and split the case where $p_{<tk} > p^*_<$ and $p_{<tk} \leq p^*_<$.

$$\sum_{k=1}^{K}\sqrt{\hat{p}_{<tk}}\hat{\sigma}_{<tk} \leq \sum_{p_{<tk}>p^*_<}^{K}\sqrt{p_{<tk}}\sigma_{<tk}\left(1 + O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)\right) + \sum_{p_{<tk}\leq p^*_<}\sqrt{p_{<tk}}\sigma_{<tk} + O\left(\frac{1}{\sqrt{N_{<t}}}\right) \tag{118}$$

$$\leq \sum_{k=1}^{K}\sqrt{p_{<tk}}\sigma_{<tk} + O\left(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\right) \tag{119}$$

Thus, for $p_{<tk} > p^*_< = \frac{2\ln(1/\delta)+2\sqrt{\ln(1/\delta)}+2}{|X^+_{<tk}|}$:

$$\hat{a}_{tk} + \frac{N_1}{N_2 K} \geq \frac{|\mathcal{D}_{<tk}|\sqrt{p_{<tk}}\sigma_{<tk} - O\Big(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\Big)}{\frac{N_2}{N}\sum_{j=1}^{K}|\mathcal{D}_{<tj}|\sqrt{p_{<tj}}\sigma_{<tj} + O\Big(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\Big)} \tag{120}$$

$$\geq \frac{(a_{tk} + \frac{N_1}{N_2 K}) - O\Big(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\Big)}{1 + O\Big(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\Big)} \tag{121}$$

$$\geq \Big((a_{tk} + \frac{N_1}{N_2 K}) - O\Big(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\Big)\Big)\Big(1 - O\Big(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\Big)\Big) \tag{122}$$

$$\geq \Big(a_{tk} + \frac{N_1}{N_2 K}\Big) - O\Big(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\Big) - O\Big(\sqrt{\frac{p_{<tk}\ln(1/\delta)}{N_{<t}}}\Big) + O\Big(\frac{\ln(1/\delta)}{N_{<t}}\Big) \tag{123}$$

$$\geq \Big(a_{tk} + \frac{N_1}{N_2 K}\Big) - O\Big(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\Big) \tag{124}$$

Subtracting the constant term from both sides and substituting $N_{<t} = (t-1)N_1/K$ gives us:

$$\hat{a}_{tk} \geq a_{tk} - O\Big(\sqrt{\frac{\ln(1/\delta)}{(t-1)N_1}}\Big) \tag{125}$$

Finally, pulling out the factor of a_{tk} and substituting $p_{<tk} = p_{tk}$ we have:

$$\hat{a}_{tk} \geq a_{tk}\Big(1 - O\Big(\sqrt{\frac{\ln(1/\delta)}{p_{tk}(t-1)N_1}}\Big)\Big) \tag{126}$$

Lemma 6.10 ($|X^+_{tk}|$ Lower Bound). *If Lemma 6.9 holds, the following is true for all k where $p_{<tk} > p^*_< = \frac{2\ln(1/\delta)+2\sqrt{\ln(1/\delta)}+2}{|X^+_{<tk}|}$ with probability at least $1-\gamma$,*

$$|X^+_{tk}| \geq p_{tk}a_{tk}N_2\Big(1 - O\Big(\sqrt{\frac{1}{p_{tk}(t-1)N_1}}\Big) - O\Big(\sqrt{\frac{\ln(1/\gamma)}{p_{tk}^{3/2}N_2}}\Big)\Big) \tag{127}$$

Proof. Recall that $|X^+_{tk}| \sim Binomial(\frac{N_1}{K} + \hat{a}_{tk}N_2)$. For simplicity, let us ignore the defensive samples – which will not invalidate the lower bound we derive – and instead use $|X^+_{tk}| \sim Binomial(\hat{a}_{tk}N_2)$. Thus, using Tarjan [4] reproduced as Lemma 7.2 we have that:

$$P\Big(|X^+_{tk}| \leq (1-\epsilon)p_{tk}\hat{a}_{tk}N_2\Big) \leq exp\Big(\frac{-\epsilon^2 p_{tk}\hat{a}_{tk}N_2}{2}\Big) \tag{128}$$

We set $\gamma = exp\Big(\frac{-\epsilon^2 p_{tk}\hat{a}_{tk}N_2}{2}\Big)$ and $\epsilon = \sqrt{\frac{2\ln(1/\gamma)}{p_{tk}\hat{a}_{tk}N_2}}$. Thus, with probability at least $1-\gamma$ and if Lemma 6.9 holds:

$$|X^+_{tk}| \geq p_{tk}\hat{a}_{tk}N_2\Big(1 - \sqrt{\frac{2\ln(1/\gamma)}{p_{tk}\hat{a}_{tk}N_2}}\Big) \tag{129}$$

We lower bound the number of draws $\hat{a}_{tk}N_2$ by conditioning on Lemma 6.9 which says that $\hat{a}_{tk} \geq a_{tk}\Big(1 - O\Big(\sqrt{\frac{\ln(1/\delta)}{p_{tk}N_{<t}}}\Big)\Big)$. Thus, with probability at least $1-\gamma$:

$$\geq p_{tk}a_{tk}N_2\left(1-O\left(\sqrt{\frac{\ln(1/\delta)}{p_{tk}N_{<t}}}\right)\right)\left(1-\sqrt{\frac{2\ln(1/\gamma)}{p_{tk}a_{tk}N_2\left(1-O\left(\sqrt{\frac{\ln(1/\delta)}{p_{tk}N_{<t}}}\right)\right)}}\right) \tag{130}$$

$$\geq p_{tk}a_{tk}N_2\left(1-O\left(\sqrt{\frac{\ln(1/\delta)}{p_{tk}N_{<t}}}\right)-\sqrt{\frac{2\ln(1/\gamma)}{p_{tk}a_{tk}N_2\left(1-O\left(\sqrt{\frac{\ln(1/\delta)}{p_{tk}N_{<t}}}\right)\right)}}\right) \tag{131}$$

$$\geq p_{tk}a_{tk}N_2\left(1-O\left(\sqrt{\frac{\ln(1/\delta)}{p_{tk}N_{<t}}}\right)-\sqrt{\frac{2\ln(1/\gamma)}{p_{tk}a_{tk}N_2}\cdot\left(1+O\left(\sqrt{\frac{\ln(1/\delta)}{p_{tk}N_{<t}}}\right)\right)}\right) \tag{132}$$

$$\geq p_{tk}a_{tk}N_2\left(1-O\left(\sqrt{\frac{\ln(1/\delta)}{p_{tk}N_{<t}}}\right)-\sqrt{\frac{2\ln(1/\gamma)}{p_{tk}a_{tk}N_2}+O\left(\sqrt{\frac{\ln(1/\delta)\ln(1/\gamma)}{p_{tk}^3N_2^2p_{tk}N_{<t}}}\right)}\right) \tag{133}$$

Finally, substituting $N_{<t}=(t-1)N_1/K$ we have:

$$|X_{tk}^+|\geq p_{tk}a_{tk}N_2\left(1-O\left(\sqrt{\frac{1}{p_{tk}(t-1)N_1}}\right)-O\left(\sqrt{\frac{\ln(1/\gamma)}{p_{tk}^{3/2}N_2}}\right)\right) \tag{134}$$

Lemma 6.11 ($w_{tk}^2P(|X_{tk}^+|=0)$ Upper Bound). *If Lemma 6.9 holds for $p_{<tk}>p_<^*$ and if $N_2=w(N_1^{3/4})$, then*

$$p_{tk}^xP(|X_{tk}^+|=0)\leq O\left(\frac{1}{N_1^x}\right)+O\left(\frac{\sqrt{N_1}}{N_2^2}\right) \tag{135}$$

Proof.

$$p_{tk}^xP(|X_{tk}^+|=0)=p_{tk}^x(1-p_{tk})^{\lceil\frac{N_1}{K}+\hat{a}_{tk}N_2\rceil}\leq p_{tk}^x(1-p_{tk})^{(\frac{N_1}{K}+\hat{a}_{tk}N_2)} \tag{136}$$

If $p_{tk}\leq p^*=\frac{2\ln(1/\delta)+2\sqrt{\ln(1/\delta)}+2}{N_1}$,

$$p_{tk}^x(1-p_{tk})^{(\frac{N_1}{K}+\hat{a}_{tk}N_2)}\leq O\left(\frac{1}{N_1^x}\right) \tag{137}$$

Else if $p_{tk}>p^*$, and the conditions of Lemma 6.9 hold, then

$$p_{tk}^x(1-p_{tk})^{(\frac{N_1}{K}+\hat{a}_{tk}N_2)}\leq(1-p_{tk})^{\left(\frac{N_1}{K}+a_{tk}N_2\left(1-O\left(\frac{1}{\sqrt{p_{tk}(t-1)N_1}}\right)\right)\right)} \tag{138}$$

$$\leq(1-p_{tk})^{\left(\frac{N_1}{K}+N_2\left(O(\sqrt{p_{tk}})-O\left(\frac{1}{\sqrt{(t-1)N_1}}\right)\right)\right)} \tag{139}$$

$$\tag{140}$$

Note that because $p_{tk}>p^*$ we have $p_{tk}>O(\frac{1}{N_1})$, thus:

$$p_{tk}^x(1-p_{tk})^{(\frac{N_1}{K}+\hat{a}_{tk}N_2)}\leq(1-p_{tk})^{\left(\frac{N_1}{K}+N_2\left(O(\frac{1}{\sqrt{N_1}})-O\left(\frac{1}{\sqrt{(t-1)N_1}}\right)\right)\right)} \tag{141}$$

$$\leq(1-p_{tk})^{\left(\frac{N_1}{K}+O\left(\frac{N_2}{N_1^{1/2}}\right)\right)} \tag{142}$$

$$\leq(1-p_{tk})^{O\left(\frac{N_2}{N_1^{1/2}}\right)} \tag{143}$$

We assume that $N_2 = w(N_1^{1/2})$ which means that (143) is decreasing exponentially. Hence we can say that $(1-p_{tk})^{O\left(\frac{N_2}{N_1^{1/2}}\right)} = O\left(\frac{N_1}{N_2^2}\right)$:

$$p_{tk}^x P(|X_{tk}^+| = 0) \leq O\left(\frac{1}{N_1^x}\right) + O\left(\frac{N_1}{N_2^2}\right) \tag{144}$$

6.2 Final Result

By Lemmas 5.1, 5.1, ..., 6.11 we have that for small $\delta > 0$:

$$\sqrt{\hat{p}_{<tk}} \leq \sqrt{p_{<tk}} \cdot \sqrt{1 + O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)} \tag{145}$$

$$\sqrt{\hat{p}_{<tk}} \geq \sqrt{p_{<tk}} \cdot \sqrt{1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)} \tag{146}$$

$$|X_{<tk}^+| \geq p_{<tk}N_{<t} - \sqrt{2\ln(1/\delta)p_{<tk}N_{<t}} \tag{147}$$

$$\sigma_{<tk}\sqrt{1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)} \leq \hat{\sigma}_{<tk} \leq \sigma_{<tk}\sqrt{1 + O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)} \tag{148}$$

$$\sqrt{\hat{p}_{tk}} \leq \sqrt{p_{tk}} \cdot \sqrt{1 + O\left(\sqrt{\frac{\ln(1/\delta)}{p_{tk}N_1}}\right)} \tag{149}$$

$$\sqrt{\hat{p}_{tk}} \geq \sqrt{p_{tk}} \cdot \sqrt{1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{tk}N_1}}\right)} \tag{150}$$

$$w_{tk} - O\left(\frac{\ln(1/\delta)\sqrt{p_{tk}}}{\sqrt{N_1}}\right) \leq \hat{w}_{tk} \leq w_{tk} + O\left(\frac{\ln(1/\delta)\sqrt{p_{tk}}}{\sqrt{N_1}}\right) \tag{151}$$

$$\sqrt{\hat{p}_{<tk}}\hat{\sigma}_{<tk} \leq \sqrt{p_{<tk}}\sigma_{<tk} + O\left(\frac{1}{\sqrt{N_{<t}}}\right) \tag{152}$$

$$\hat{a}_{tk} \geq a_{tk}\left(1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{tk}(t-1)N_1}}\right)\right) \tag{153}$$

$$|X_{tk}^+| \geq p_{tk}a_{tk}N_2\left(1 - O\left(\sqrt{\frac{1}{p_{tk}(t-1)N_1}}\right) - O\left(\sqrt{\frac{\ln(1/\gamma)}{p_{tk}^{3/2}N_2}}\right)\right) \tag{154}$$

$$p_{tk}^x P(|X_{tk}^+| = 0) \leq O\left(\frac{1}{N_1^x}\right) + O\left(\frac{\sqrt{N_1}}{N_2^2}\right) \tag{155}$$

Let $\mathcal{E}$ be the event that all these inequalities are satisfied. For the remainder of the proof we condition on $\mathcal{E}$ and take the expectation of the randomness of our draws in segments $1, \ldots, t$.

We'll begin by decomposing the mean squared error:

$$\mathbb{E}[(\hat{\mu}_t - \mu_t)^2] = \mathbb{E}\Big[\Big(\sum_{k=1}^{K} \hat{w}_{tk}\hat{\mu}_{tk} - w_{tk}\mu_{tk}\Big)^2\Big] \tag{156}$$

We know that $\mathbb{E}[\hat{\mu}_{tk}] = P(|X_{tk}^+| > 0) \cdot \mu_{tk}$. As a result, $\mathbb{E}[\hat{\mu}_{tk}] + P(|X_{tk}^+| = 0) \cdot \mu_{tk} = \mu_{tk}$. Hence,

$$= \mathbb{E}\Big[\Big(\sum_{k=1}^{K} \hat{w}_{tk}(\hat{\mu}_{tk} - \mathbb{E}[\hat{\mu}_{tk}]) + (\hat{w}_{tk} - w_{tk})\mathbb{E}[\hat{\mu}_{tk}] - w_{tk}\mu_{tk}P(|X_{tk}^+| = 0)\Big)^2\Big] \tag{157}$$

We now use that $\mathbb{E}[(A+B)^2] = \mathbb{E}[A^2 + 2AB + B^2]$, so if $\mathbb{E}[A] = 0$ then $\mathbb{E}[(A+B)^2] = \mathbb{E}[A^2] + \mathbb{E}[B^2]$. We set $A = \hat{w}_{tk}(\hat{\mu}_{tk} - \mathbb{E}[\hat{\mu}_{tk}])$ and $B = (\hat{w}_{tk} - w_{tk})\mathbb{E}[\hat{\mu}_{tk}] - w_{tk}\mu_{tk}P(|X_{tk}^+| = 0)$. Note that $\mathbb{E}[A] = 0$. Thus,

$$= \mathbb{E}\Big[\Big(\sum_{k=1}^{K} \hat{w}_{tk}(\hat{\mu}_{tk} - \mathbb{E}[\hat{\mu}_{tk}])\Big)^2\Big] + \mathbb{E}\Big[\Big((\hat{w}_{tk} - w_{tk})\mathbb{E}[\hat{\mu}_{tk}] - w_{tk}\mu_{tk}P(|X_{tk}^+| = 0)\Big)^2\Big] \tag{158}$$

$$\leq \sum_{k=1}^{K} \hat{w}_{tk}^2 \mathrm{Var}[\hat{\mu}_{tk}] + \Big[\max_k C_k^{\mu_{tk}^2}\Big]\mathbb{E}\Big[\Big(w_{tk}P(|X_{tk}^+| = 0) + (\hat{w}_{tk} - w_{tk})\Big)^2\Big] \tag{159}$$

We will now separately upper bound the two terms in this expression. For the first term, notice that:

$$\begin{aligned}
\mathrm{Var}[\hat{\mu}_{tk}] &= \mathbb{E}[\hat{\mu}_{tk}^2] - \mathbb{E}[\hat{\mu}_{tk}]^2 && (160)\\
&= P(|X_{tk}^+| > 0)(\mathbb{E}\Big[\frac{\sigma_{tk}^2}{|X_{tk}^+|}\Big| |X_{tk}^+| > 0\Big] + \mu_{tk}^2) - P(|X_{tk}^+| > 0)^2\mu_{tk}^2 && (161)\\
&= P(|X_{tk}^+| > 0)\mathbb{E}\Big[\frac{\sigma_{tk}^2}{|X_{tk}^+|}\Big| |X_{tk}^+| > 0\Big] + P(|X_{tk}^+| > 0)(1 - P(|X_{tk}^+| > 0))\mu_{tk}^2 && (162)\\
&= P(|X_{tk}^+| > 0)\mathbb{E}\Big[\frac{\sigma_{tk}^2}{|X_{tk}^+|}\Big| |X_{tk}^+| > 0\Big] + P(|X_{tk}^+| > 0)P(|X_{tk}^+| = 0)\mu_{tk}^2 && (163)\\
&\leq \mathbb{E}\Big[\frac{\sigma_{tk}^2}{|X_{tk}^+|}\Big| |X_{tk}^+| > 0\Big] + P(|X_{tk}^+| = 0)\mu_{tk}^2 && (164)
\end{aligned}$$

$$\begin{aligned}
\sum_{k=1}^{K} \hat{w}_{tk}^2 \mathrm{Var}[\hat{\mu}_{tk}] &\leq \sum_{k=1}^{K} \hat{w}_{tk}^2 \Big(\mathbb{E}\Big[\frac{\sigma_{tk}^2}{|X_{tk}^+|}\Big| |X_{tk}^+| > 0\Big] + P(|X_{tk}^+| = 0)\mu_{tk}^2\Big) && (165)\\
&\leq \sum_{k=1}^{K} \Bigg(O\Big(p_{tk}^2\Big) + O\Big(\frac{p_{tk}^{3/2}}{\sqrt{N_1}}\Big)\Bigg)\Big(\mathbb{E}\Big[\frac{\sigma_{tk}^2}{|X_{tk}^+|}\Big| |X_{tk}^+| > 0\Big] + P(|X_{tk}^+| = 0)\mu_{tk}^2\Big) && (166)
\end{aligned}$$

We split the cases where $p_{<tk}$ is small and $p_{<tk}$ is large via $p_<^* = \frac{2\ln(1/\delta) + 2\sqrt{\ln(1/\delta)} + 2}{|X_{<tk}^+|}$. This value of $p_<^*$ is chosen as according to the conditions of Lemma 6.9.

$$\leq \sum_{p_{<tk} > p_<^*}^{K} \Bigg(O\big(p_{tk}^2\big) + O\Big(\frac{p_{tk}^{3/2}}{\sqrt{N_1}}\Big)\Bigg)\Big(\mathbb{E}\Big[\frac{\sigma_{tk}^2}{|X_{tk}^+|}\Big| |X_{tk}^+| > 0\Big] + P(|X_{tk}^+| = 0)\mu_{tk}^2\Big) + \sum_{p_{<tk} \leq p_<^*}^{K} O\Bigg(\frac{1}{\sqrt{N_{<t}}}\Bigg) \tag{167}$$

Note that $p_{tk}^2 P(|X_{tk}^+| = 0) \leq O\big(\frac{1}{N_1^2}\big) + O\big(\frac{N_1}{N_2^2}\big)$ According to Lemma 6.11

$$\leq \sum_{p_{<tk} > p_<^*}^{K} \Bigg(O\big(p_{tk}^2\big) + O\Big(\frac{p_{tk}^{3/2}}{\sqrt{N_1}}\Big)\Bigg)\mathbb{E}\Big[\frac{\sigma_{tk}^2}{|X_{tk}^+|}\Big| |X_{tk}^+| > 0\Big] + O\Big(\frac{1}{N_1^2}\Big) + O\Big(\frac{N_1}{N_2^2}\Big) \tag{168}$$

We will now apply Lemma 6.10 to upper bound $\mathbb{E}\Big[\frac{\sigma_{tk}^2}{|X_{tk}^+|}\Big||X_{tk}^+| > 0\Big]$. Namely, we split the expectation into two cases, one where $|X_{tk}^+| \geq F_{tk}$ and one where $|X_{tk}^+| < F_{tk}$. Specifically, we apply it such that $|X_{tk}^+| \geq F_{tk} = p_{tk}a_{tk}N_2\Bigg(1-O\bigg(\sqrt{\frac{1}{p_{tk}(t-1)N_1}}\bigg)-O\bigg(\sqrt{\frac{\ln(1/\gamma)}{p_{tk}^{3/2}N_2}}\bigg)\Bigg)$ with failure probability $\gamma = \frac{1}{e^{\sqrt{N_1}}} \leq O\big(\frac{1}{N_1}\big)$.

$$\leq \sum_{p_{<tk}>p_<^*}^{K} \Bigg(O(p_{tk}^2) + O\bigg(\frac{p_{tk}^{3/2}}{\sqrt{N_1}}\bigg)\Bigg)\Bigg(\mathbb{E}\Big[\frac{\sigma_{tk}^2}{|X_{tk}^+|}\Big||X_{tk}^+| > 0\Big]\Bigg) + O\bigg(\frac{1}{N_1^2}\bigg) + O\bigg(\frac{N_1}{N_2^2}\bigg) \tag{169}$$

$$\leq \sum_{p_{<tk}>p_<^*}^{K} \Bigg(\dots\Bigg)\Bigg(P(|X_{tk}^+| > F_{tk})\mathbb{E}\Big[\frac{\sigma_{tk}^2}{|X_{tk}^+|}\Big||X_{tk}^+| > 0\Big] + O\bigg(\frac{1}{N_1}\bigg)C_{tk}^{\sigma_{tk}^2}\Bigg) + O\bigg(\frac{1}{N_1^2}\bigg) + O\bigg(\frac{N_1}{N_2^2}\bigg) \tag{170}$$

$$\leq \sum_{p_{<tk}>p_<^*}^{K} \Bigg(O(p_{tk}^2) + O\bigg(\frac{p_{tk}^{3/2}}{\sqrt{N_1}}\bigg)\Bigg)\Bigg(\mathbb{E}\Big[\frac{\sigma_{tk}^2}{|X_{tk}^+|}\Big||X_{tk}^+| > 0\Big] + O\bigg(\frac{1}{N_1}\bigg)C_{tk}^{\sigma_{tk}^2}\Bigg) + O\bigg(\frac{1}{N_1^2}\bigg) + O\bigg(\frac{N_1}{N_2^2}\bigg) \tag{171}$$

$$\leq \sum_{p_{<tk}>p_<^*}^{K} \Bigg(O(p_{tk}^2) + O\bigg(\frac{p_{tk}^{3/2}}{\sqrt{N_1}}\bigg)\Bigg)\mathbb{E}\Big[\frac{\sigma_{tk}^2}{|X_{tk}^+|}\Big||X_{tk}^+| > 0\Big] + O\bigg(\frac{1}{N_1^2}\bigg) + O\bigg(\frac{N_1}{N_2^2}\bigg) \tag{172}$$

$$\leq \sum_{p_{<tk}>p_<^*}^{K} \Bigg(O(p_{tk}^2) + O\bigg(\frac{p_{tk}^{3/2}}{\sqrt{N_1}}\bigg)\Bigg)\frac{\sigma_{tk}^2}{p_{tk}a_{tk}N_2\Bigg(1 - O\bigg(\sqrt{\frac{1}{p_{tk}(t-1)N_1}}\bigg) - O\bigg(\sqrt{\frac{\ln(1/\gamma)}{p_{tk}^{3/2}N_2}}\bigg)\Bigg)} + O\bigg(\frac{1}{N_1^2}\bigg) + O\bigg(\frac{N_1}{N_2^2}\bigg) \tag{173}$$

$$\leq \sum_{p_{<tk}>p_<^*}^{K} \Bigg(O\bigg(\frac{\sqrt{p_{tk}}}{N_2}\bigg) + O\bigg(\frac{1}{N_2\sqrt{N_1}}\bigg)\Bigg)\Bigg(1 + O\bigg(\sqrt{\frac{1}{p_{tk}(t-1)N_1}}\bigg) + O\bigg(\sqrt{\frac{\ln(1/\gamma)}{p_{tk}^{3/2}N_2}}\bigg)\Bigg) + O\bigg(\frac{1}{N_1^2}\bigg) + O\bigg(\frac{N_1}{N_2^2}\bigg) \tag{174}$$

Since we condition on $p_{<tk} > p_<^*$ and we assume $p_{tk} = p_{<tk}$, we can substitute $O(1/N_1)$ for p_{tk} and $\ln(1/\gamma) = \sqrt{N_1}$:

$$\leq \sum_{p_{<tk}>p_<^*}^{K} \Bigg(O\bigg(\frac{1}{N_2\sqrt{N_1}}\bigg) + O\bigg(\frac{1}{N_2\sqrt{N_1}}\bigg)\Bigg)\Bigg(1 + O\bigg(\sqrt{\frac{N_1}{(t-1)N_1}}\bigg) + O\bigg(\sqrt{\frac{N_1^2}{N_2}}\bigg)\Bigg) + O\bigg(\frac{1}{N_1^2}\bigg) + O\bigg(\frac{N_1}{N_2^2}\bigg) \tag{175}$$

$$\leq \sum_{p_{<tk}>p_<^*}^{K} O\bigg(\frac{1}{N_2\sqrt{N_1}}\bigg)\Bigg(1 + O\bigg(\sqrt{\frac{N_1}{(t-1)N_1}}\bigg) + O\bigg(\sqrt{\frac{N_1^2}{N_2}}\bigg)\Bigg) + O\bigg(\frac{1}{N_1^2}\bigg) + O\bigg(\frac{N_1}{N_2^2}\bigg) \tag{176}$$

$$\leq \sum_{p_{<tk}>p_<^*}^{K} O\bigg(\frac{1}{N_2\sqrt{N_1}}\bigg) + O\bigg(\frac{1}{N_2\sqrt{(t-1)N_1}}\bigg) + O\bigg(\frac{\sqrt{N_1}}{N_2\sqrt{N_2}}\bigg) + O\bigg(\frac{1}{N_1^2}\bigg) + O\bigg(\frac{N_1}{N_2^2}\bigg) \tag{177}$$

$$\leq O\bigg(\frac{1}{N_2\sqrt{N_1}}\bigg) + O\bigg(\frac{1}{N_2\sqrt{N_1 t}}\bigg) + O\bigg(\frac{\sqrt{N_1}}{N_2\sqrt{N_2}}\bigg) + O\bigg(\frac{1}{N_1^2}\bigg) + O\bigg(\frac{N_1}{N_2^2}\bigg) \tag{178}$$

We will now upper bound the second term in the expression on (159). We apply Lemma 6.11 on (179):

$$\left[\max_k C_k^{\mu_{tk}^2}\right]\mathbb{E}\left[\left(w_{tk}P(|X_{tk}^+|=0)+(\hat{w}_{tk}-w_{tk})\right)^2\right] \leq \left[\max_k C_k^{\mu_{tk}^2}\right]\mathbb{E}\left[\left(O\left(\frac{1}{N_1}\right)+O\left(\frac{N_1}{N_2^2}\right)+O\left(\frac{1}{\sqrt{N_1}}\right)\right)^2\right] \tag{179}$$

$$\leq \left[\max_k C_k^{\mu_{tk}^2}\right]\mathbb{E}\left[\left(O\left(\frac{N_1}{N_2^2}\right)+O\left(\frac{1}{\sqrt{N_1}}\right)\right)^2\right] \tag{180}$$

$$\leq O\left(\frac{N_1^2}{N_2^4}\right)+O\left(\frac{\sqrt{N_1}}{N_2^2}\right)+O\left(\frac{1}{N_1}\right) \tag{181}$$

Finally, putting (178) and (181) together we have:

$$\mathbb{E}[(\hat{\mu}_t-\mu_t)^2] = \mathbb{E}\left[\left(\sum_{k=1}^{K}\hat{w}_{tk}\hat{\mu}_{tk}-w_{tk}\mu_{tk}\right)^2\right] \tag{182}$$

$$= \sum_{k=1}^{K}\hat{w}_{tk}^2\mathrm{Var}[\hat{\mu}_{tk}]+\left[\max_k C_k^{\mu_{tk}^2}\right]\mathbb{E}\left[\left(w_{tk}P(|X_{tk}^+|=0)+(\hat{w}_{tk}-w_{tk})\right)^2\right] \tag{183}$$

$$\leq O\left(\frac{1}{N_2\sqrt{N_1}}\right)+O\left(\frac{1}{N_2\sqrt{N_1t}}\right)+O\left(\frac{\sqrt{N_1}}{N_2\sqrt{N_2}}\right)+O\left(\frac{1}{N_1^2}\right)+O\left(\frac{N_1}{N_2^2}\right)+O\left(\frac{N_1^2}{N_2^4}\right)+O\left(\frac{\sqrt{N_1}}{N_2^2}\right)+O\left(\frac{1}{N_1}\right) \tag{184}$$

$$\leq O\left(\frac{1}{N_2\sqrt{N_1}}\right)+O\left(\frac{1}{N_2\sqrt{N_1t}}\right)+O\left(\frac{\sqrt{N_1}}{N_2\sqrt{N_2}}\right)+O\left(\frac{N_1}{N_2^2}\right)+O\left(\frac{N_1^2}{N_2^4}\right)+O\left(\frac{1}{N_1}\right) \tag{185}$$

$$\leq O\left(\frac{1}{N_1}\right)+O\left(\frac{1}{N_2\sqrt{N_1t}}\right)+O\left(\frac{N_1}{N_2^2}\right)+O\left(\frac{1}{N_2\sqrt{N_1}}\right) \tag{186}$$

7 Appendix

7.1 Lemmas

Lemma 7.1 (Chung and Lu [1]) *Let $X_1,\ldots,X_n$ be independent random variables with:*

$$P(X_i=1)=p_i, P(X_i=0)=1-p_i \tag{187}$$

For $X=\sum_{i=1}^{N}a_iX_i$, we have $\mathbb{E}[X]=\sum_{i=1}^{N}a_ip_i$ and if we define $v=\sum_{i=1}^{N}a_i^2p_i$. Then we have:

$$P(x<\mathbb{E}[X]-\lambda)\leq e^{-\lambda^2/2v} \tag{188}$$

$$P(x>\mathbb{E}[X]+\lambda)\leq e^{-\frac{\lambda^2}{2(v+a\lambda/3)}} \tag{189}$$

Lemma 7.2 (Tarjan [4]) *Let $X_1,\ldots,X_n$ be independent random variables that are not necessarily from the same distribution. Assume that $0\leq X_i\leq 1$ for each i. Let $X=X_1+\cdots+X_n$. Write $\mu=\mathbb{E}[X]=\mathbb{E}[X_1]+\cdots+\mathbb{E}[X_n]$. Then, for any $\epsilon\geq 0$:*

$$P(X\geq(1+\epsilon)\mu)\leq\exp\left(-\frac{\epsilon^2}{2+\epsilon^{\mu}}\right) \tag{190}$$

$$P(X\leq(1-\epsilon)\mu)\leq\exp\left(-\frac{\epsilon^2}{2}\mu\right) \tag{191}$$

Lemma 7.3 (Ferguson [2]) *Let $X_1,\ldots,X_n$ be i.i.d. For a U-statistic of order 2 in the form $U_n=\frac{1}{n2}\sum_{i<j}g(X_i,X_j)$, if $g(X_i,X_j)=\frac{1}{2}(X_i-X_j)^2$:*

$$\mathbb{E}[U_n] = Var[X] \tag{192}$$

Lemma 7.4 (Rinaldo [3]) *Let $X_1, \ldots, X_n$ be i.i.d. For a U-statistic of order 2 in the form $U_n = \frac{1}{n2}\sum_{i<j} g(X_i, X_j)$, with an upper bound $g(X_i, X_j) \leq b$:*

$$P(|U_n - \mathbb{E}[U_n]| \geq t) \leq 2\text{exp}\Big(-\frac{nt^2}{8b^2}\Big) \tag{193}$$

7.2 Tables

We reproduce Tables 3 and 4 from our paper with RMSE metrics reported to three significant digits:

Table 2: Summary of algorithm performance relative to streaming baselines and ABae in the no predicate case. RMSE errors are computed by taking the geometric mean of the average RMSE across all datasets at the specified budget.

Algorithm	$NT = 500$	$NT = 2500$	$NT = 5000$	All
$RMSE_{\text{uniform}^\dagger}$	6.49e-2	2.88e-2	1.97e-2	2.97e-2
$RMSE_{\text{stratified}^*}$	6.39e-2	2.86e-2	1.96e-2	2.98e-2
$RMSE_{\text{ABae}^\ddagger}$	4.43e-2	1.49e-2	1.03e-2	1.64e-2
$RMSE_{\text{InQuest}}$	**3.17e-2**	**1.41e-2**	**9.91e-3**	**1.49e-2**
Improvement†	2.05x	2.03x	1.99x	2.00x
Improvement*	2.01x	2.02x	1.98x	2.00x
Improvement‡	1.40x	1.05x	1.04x	1.10x

Table 3: Summary of algorithm performance relative to streaming baselines and ABae in predicate case. RMSE errors are computed by taking the geometric mean of the average RMSE across all datasets at the specified budget.

Algorithm	$NT = 500$	$NT = 2500$	$NT = 5000$	All
$RMSE_{\text{uniform}^\dagger}$	7.22e-2	3.16e-2	2.14e-2	3.28e-2
$RMSE_{\text{stratified}^*}$	6.45e-2	2.90e-2	2.00e-2	3.03e-2
$RMSE_{\text{ABae}^\ddagger}$	9.62e-2	2.67e-2	1.71e-2	2.92e-2
$RMSE_{\text{InQuest}}$	**4.87e-2**	**2.03e-2**	**1.36e-2**	**2.13e-2**
Improvement†	1.48x	1.56x	1.58x	1.54x
Improvement*	1.32x	1.43x	1.48x	1.42x
Improvement‡	1.97x	1.32x	1.26x	1.37x

7.3 Figures

7.3.1 Sensitivity Analysis

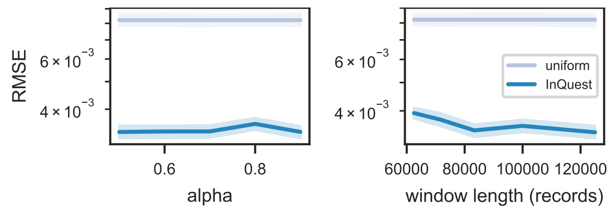

Figure 1: Sensitivity analysis of InQuest to the smoothing parameter α and our tumbling window length on the `archie` dataset on the evaluation query without a predicate.

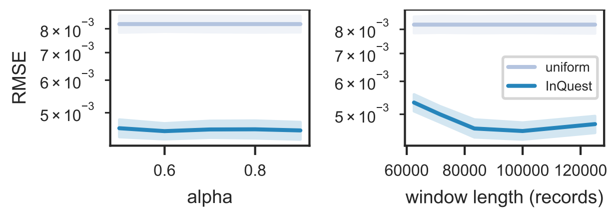

Figure 2: Sensitivity analysis of InQuest to the smoothing parameter α and our tumbling window length on the `archie` dataset on the evaluation query with a predicate.

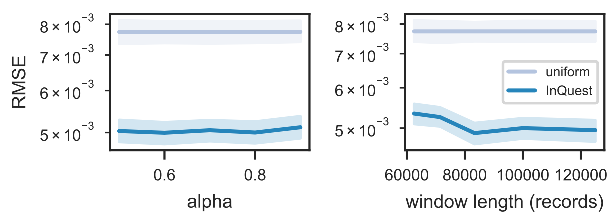

Figure 3: Sensitivity analysis of InQuest to the smoothing parameter α and our tumbling window length on the `customer-support` dataset on the evaluation query without a predicate.

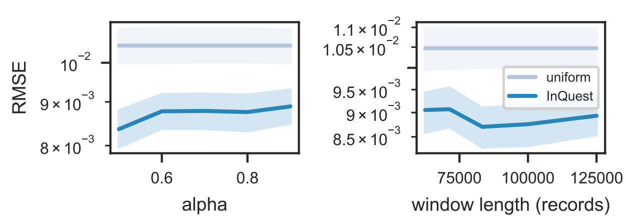

Figure 4: Sensitivity analysis of InQuest to the smoothing parameter α and our tumbling window length on the `customer-support` dataset on the evaluation query with a predicate.

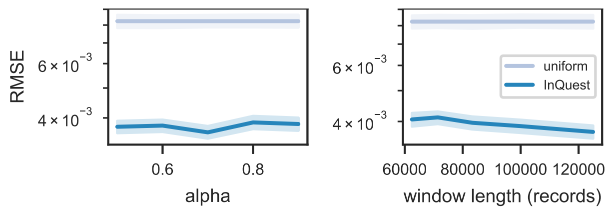

Figure 5: Sensitivity analysis of InQuest to the smoothing parameter α and our tumbling window length on the `night-street` dataset on the evaluation query without a predicate.

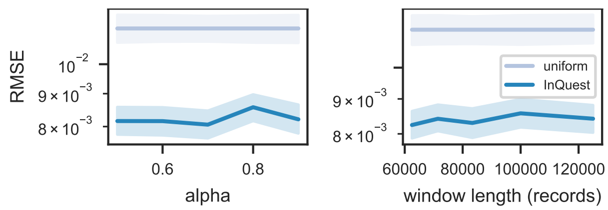

Figure 6: Sensitivity analysis of InQuest to the smoothing parameter α and our tumbling window length on the `night-street` dataset on the evaluation query with a predicate.

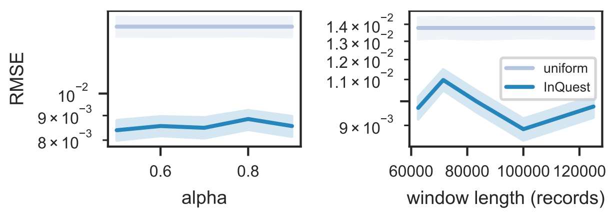

Figure 7: Sensitivity analysis of InQuest to the smoothing parameter α and our tumbling window length on the `rialto` dataset on the evaluation query without a predicate.

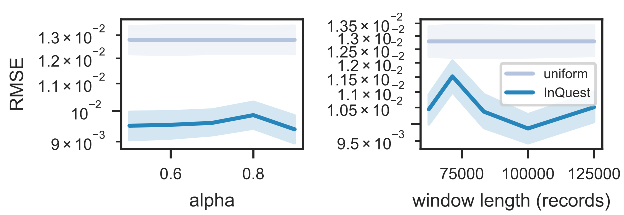

Figure 8: Sensitivity analysis of InQuest to the smoothing parameter α and our tumbling window length on the `rialto` dataset on the evaluation query with a predicate.

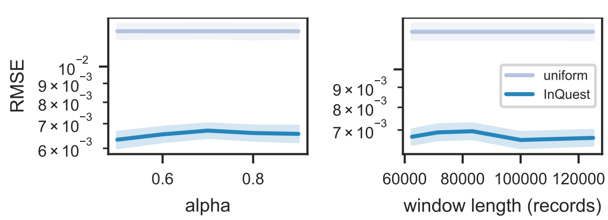

Figure 9: Sensitivity analysis of InQuest to the smoothing parameter α and our tumbling window length on the `taipei` dataset on the evaluation query without a predicate.

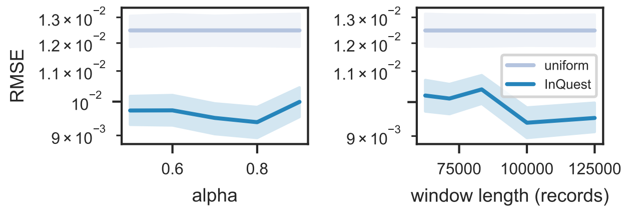

Figure 10: Sensitivity analysis of InQuest to the smoothing parameter α and our tumbling window length on the `taipei` dataset on the evaluation query with a predicate.

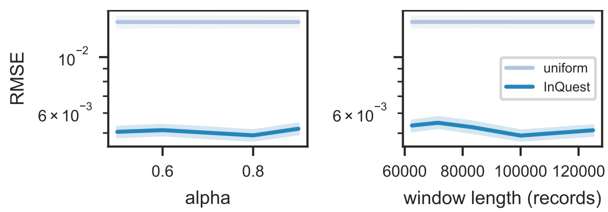

Figure 11: Sensitivity analysis of InQuest to the smoothing parameter α and our tumbling window length on the `grand-canal` dataset on the evaluation query without a predicate.

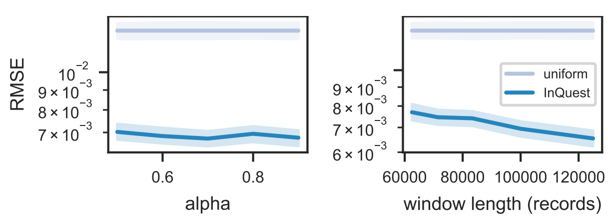

Figure 12: Sensitivity analysis of InQuest to the smoothing parameter α and our tumbling window length on the `grand-canal` dataset on the evaluation query with a predicate.

7.3.2 Proxy Quality

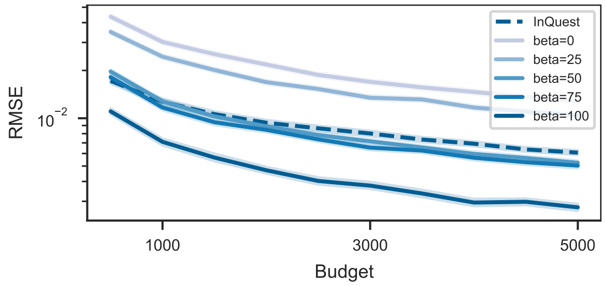

Figure 13: Proxy quality's effect on InQuest's performance on the `night-street` dataset. We plot InQuest's performance on the median segment RMSE metric as a function of β for the evaluation query without a predicate.

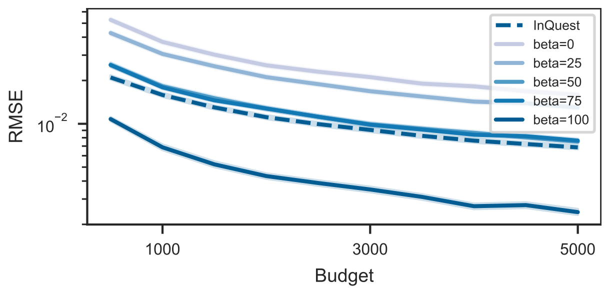

Figure 14: Proxy quality's effect on InQuest's performance on the `archie` dataset. We plot InQuest's performance on the median segment RMSE metric as a function of β for the evaluation query without a predicate.

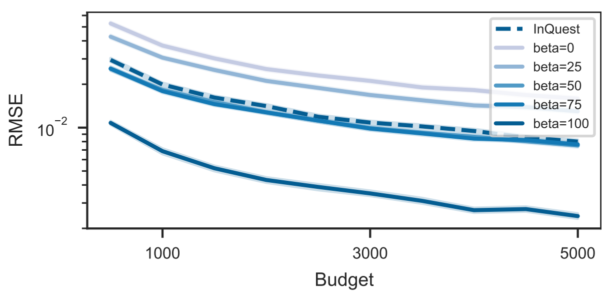

Figure 15: Proxy quality's effect on InQuest's performance on the `archie` dataset. We plot InQuest's performance on the median segment RMSE metric as a function of β for the evaluation query with a predicate.

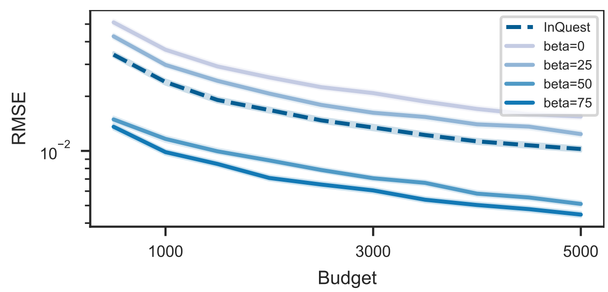

Figure 16: Proxy quality's effect on InQuest's performance on the `customer-support` dataset. We plot InQuest's performance on the median segment RMSE metric as a function of β for the evaluation query without a predicate.

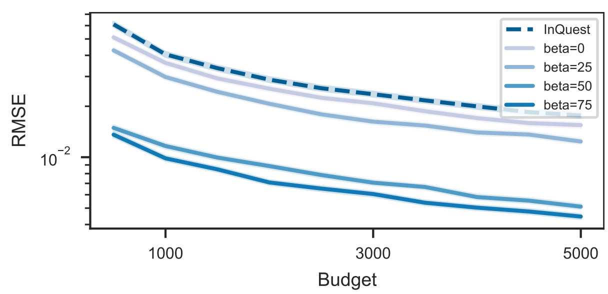

Figure 17: Proxy quality's effect on InQuest's performance on the `customer-support` dataset. We plot InQuest's performance on the median segment RMSE metric as a function of β for the evaluation query with a predicate.

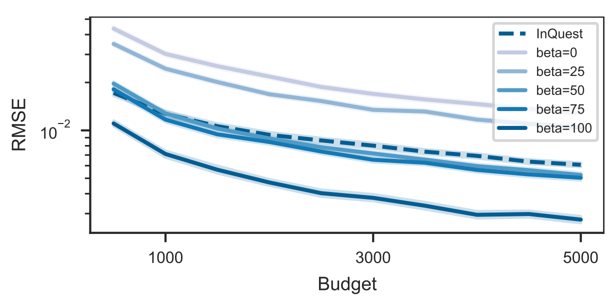

Figure 18: Proxy quality's effect on InQuest's performance on the `night-street` dataset. We plot InQuest's performance on the median segment RMSE metric as a function of β for the evaluation query without a predicate.

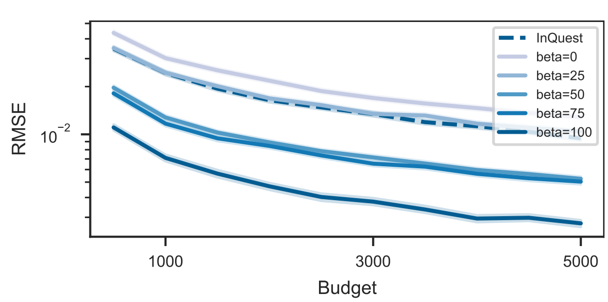

Figure 19: Proxy quality's effect on InQuest's performance on the `night-street` dataset. We plot InQuest's performance on the median segment RMSE metric as a function of β for the evaluation query with a predicate.

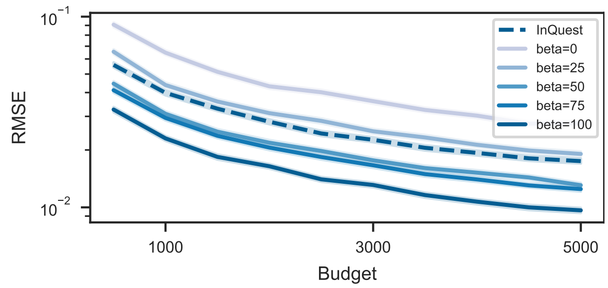

Figure 20: Proxy quality's effect on InQuest's performance on the `rialto` dataset. We plot InQuest's performance on the median segment RMSE metric as a function of β for the evaluation query without a predicate.

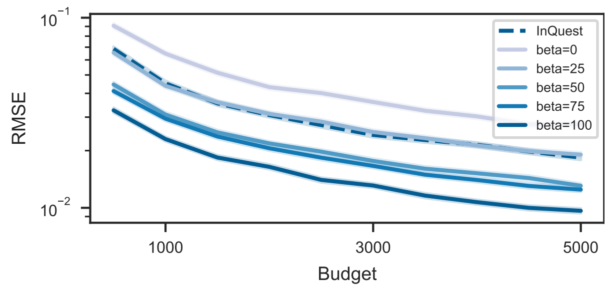

Figure 21: Proxy quality's effect on InQuest's performance on the `rialto` dataset. We plot InQuest's performance on the median segment RMSE metric as a function of β for the evaluation query with a predicate.

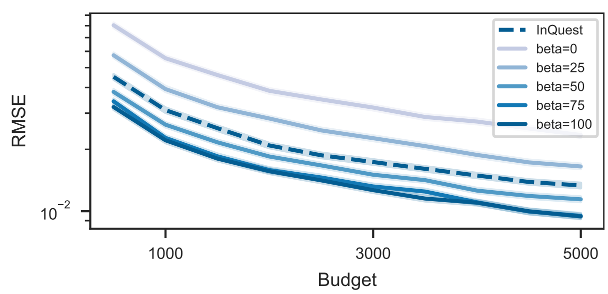

Figure 22: Proxy quality's effect on InQuest's performance on the `taipei` dataset. We plot InQuest's performance on the median segment RMSE metric as a function of β for the evaluation query without a predicate.

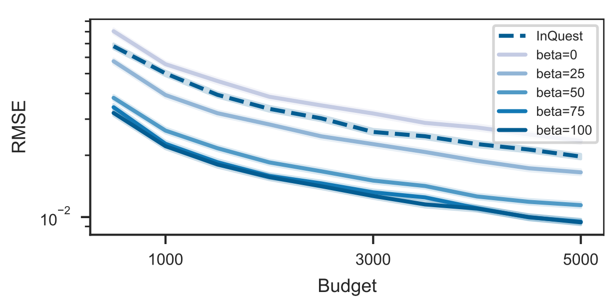

Figure 23: Proxy quality's effect on InQuest's performance on the `taipei` dataset. We plot InQuest's performance on the median segment RMSE metric as a function of β for the evaluation query with a predicate.

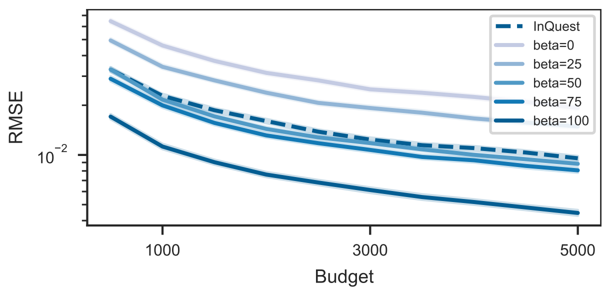

Figure 24: Proxy quality's effect on InQuest's performance on the `venice` dataset. We plot InQuest's performance on the median segment RMSE metric as a function of β for the evaluation query without a predicate.

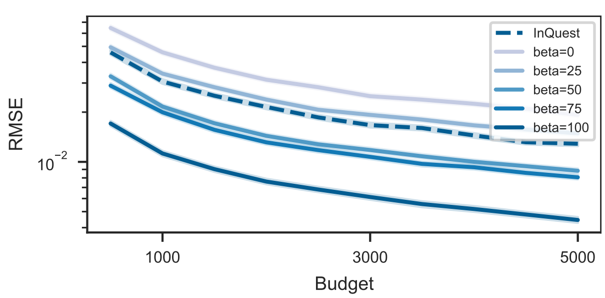

Figure 25: Proxy quality's effect on InQuest's performance on the `venice` dataset. We plot InQuest's performance on the median segment RMSE metric as a function of β for the evaluation query with a predicate.

7.3.3 Cost Savings

Cost-savings are based on estimates for running object detection DNN's on an NVIDIA T4 GPU, thus we present plots for all of our datasets except for `customer-support` (which is a text dataset).

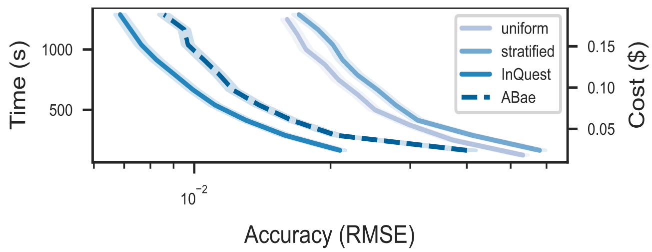

Figure 26: Time and cost in dollars as a function of accuracy for the `archie` dataset on the evaluation queries without a predicate.

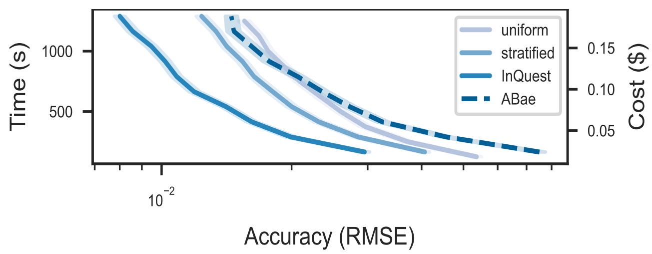

Figure 27: Time and cost in dollars as a function of accuracy for the `archie` dataset on the evaluation queries with a predicate.

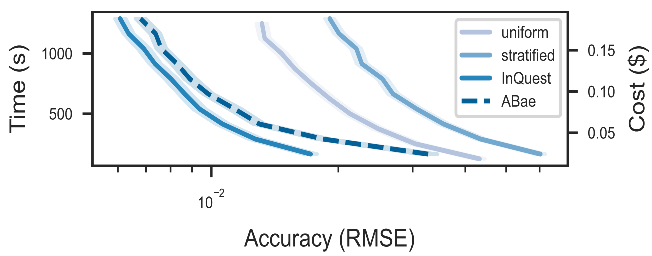

Figure 28: Time and cost in dollars as a function of accuracy for the `night-street` dataset on the evaluation queries without a predicate.

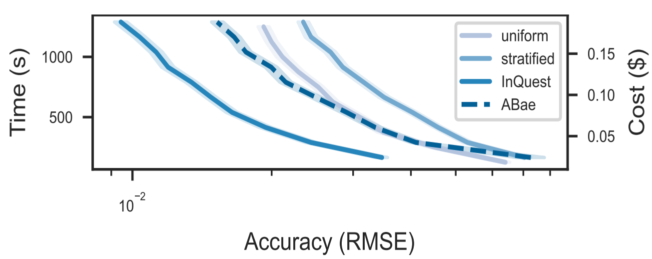

Figure 29: Time and cost in dollars as a function of accuracy for the `night-street` dataset on the evaluation queries with a predicate.

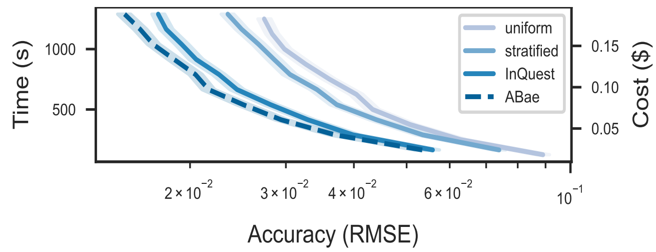

Figure 30: Time and cost in dollars as a function of accuracy for the `rialto` dataset on the evaluation queries without a predicate.

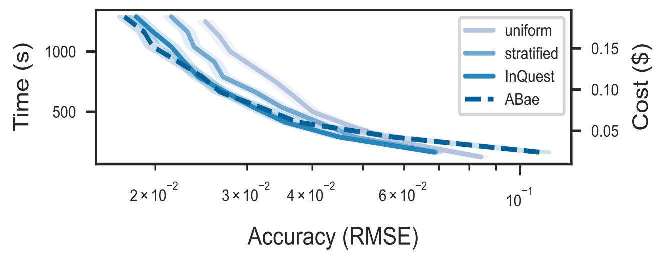

Figure 31: Time and cost in dollars as a function of accuracy for the `rialto` dataset on the evaluation queries with a predicate.

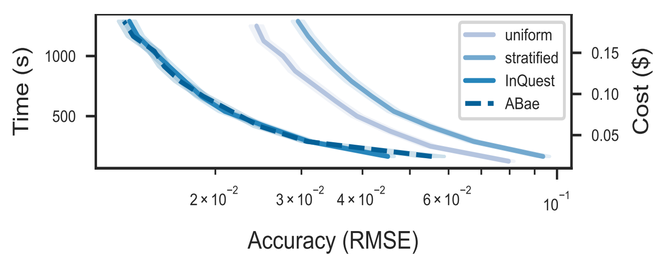

Figure 32: Time and cost in dollars as a function of accuracy for the `taipei` dataset on the evaluation queries without a predicate.

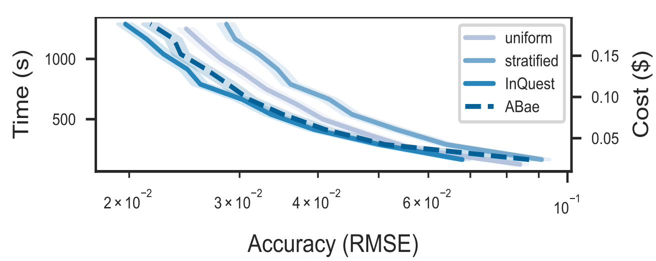

Figure 33: Time and cost in dollars as a function of accuracy for the `taipei` dataset on the evaluation queries with a predicate.

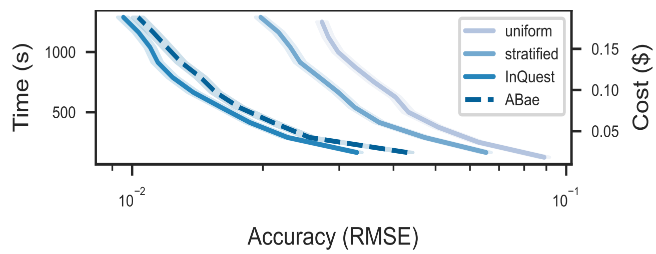

Figure 34: Time and cost in dollars as a function of accuracy for the `venice` dataset on the evaluation queries without a predicate.

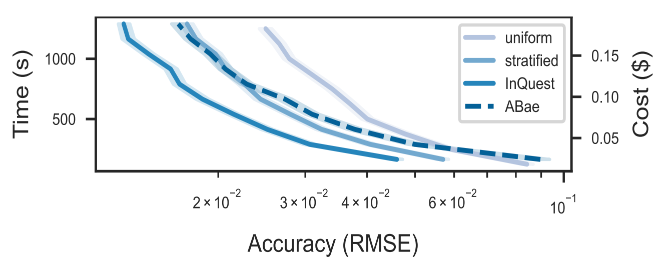

Figure 35: Time and cost in dollars as a function of accuracy for the `venice` dataset on the evaluation queries with a predicate.

7.3.4 Full Query RMSE

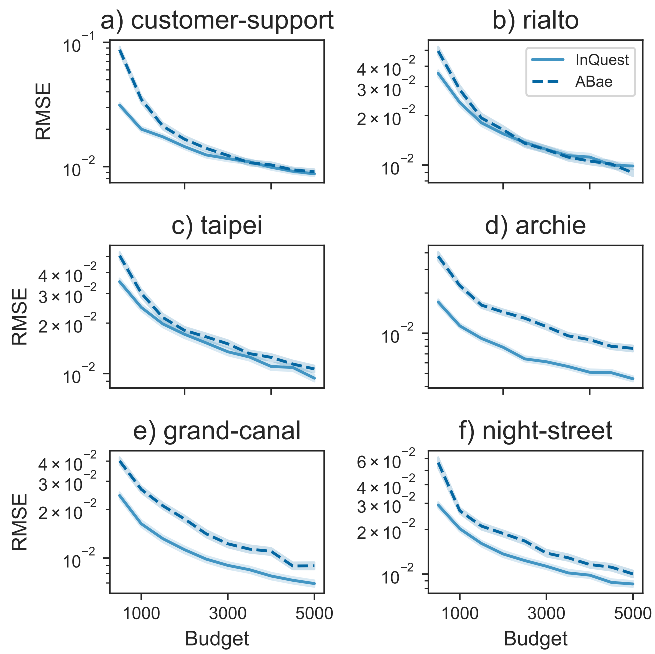

Figure 36: Sample budget vs. full query RMSE for InQuest and ABae on the evaluation queries with a predicate (log scale).

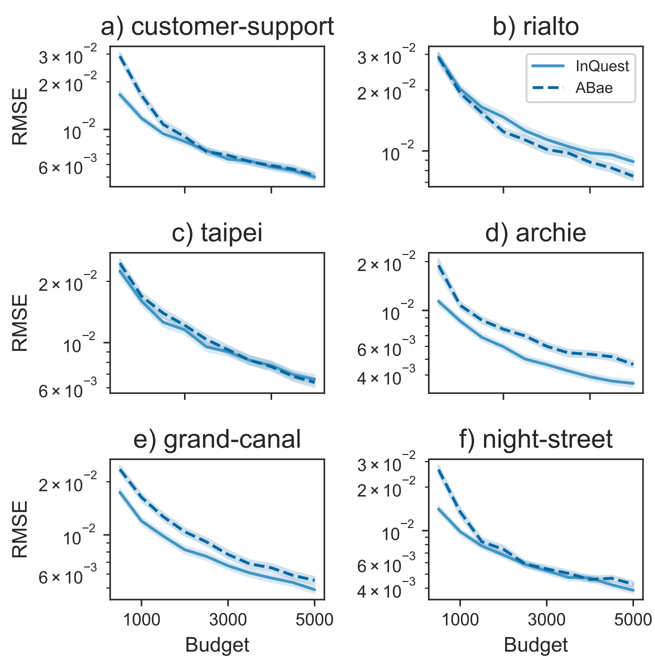

Figure 37: Sample budget vs. full query RMSE for InQuest and ABae on the evaluation queries without a predicate (log scale).

References

[1] Fan Chung and Linyuan Lu. "Connected Components in Random Graphs with Given Expected Degree Sequences." In: *Annals of Combinatorics* 6 (2 Nov. 2002), pp. 125–145. DOI: 10.1007/PL00012580. URL: https://doi.org/10.1007/PL00012580.

[2] Thomas Ferguson. *U-statistics notes for statistics 200c, spring 2005.* 2005. URL: https://www.stat.cmu.edu/~arinaldo/Teaching/36710/F18/Scribed_Lectures/Sep24.pdf.

[3] Alessandro Rinaldo. *36-710: Advanced statistical theory lecture 7: September 24.* 2018. URL: https://www.stat.cmu.edu/~arinaldo/Teaching/36710/F18/Scribed_Lectures/Sep24.pdf.

[4] Robert Tarjan. *15-359: Probability and computing lecture 10: More chernoff bounds, sampling, and the chernoff + union bound method*. 2009. URL: https://www.cs.princeton.edu/courses/archive/fall09/cos521/Handouts/probabilityandcomputing.pdf.