streaming-approx-queries-technical-report

Authors

October 2022

Abstract

TODO

Table 1: Summary of notation.

Symbol	Description
$\mathcal{D}$	Streaming dataset of records
$\mathcal{S}$	Stratification, i.e. k strata
$\mathcal{D}_{tk}$	The set of dataset records in (segment t, strata k)
$\mathcal{D}_{tk}^{+}$	The set of predicate-matching dataset records in (segment t, strata k)
$\mathcal{P}(\S)$	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
K	Number of strata
$O(x)$	Oracle predicate
$X_{tk,i}$	ith sample from (segment t, strata k)
w_{tk}	$\lvert D_{tk}\rvert p_{tk}/\sum \lvert D_{tj}\rvert p_{tj}$
S_{tk}	The set of predicate matching samples drawn from $\mathcal{D}_{tk}$
p_{tk}	Predicate positive rate
...	...

1 Setup

TODO

2 Algorithm

TODO

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:

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

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:

$$\hat{\mu}_t = \sum_{k=1}^{K} w_{tk} \cdot \frac{\sum_{x \in S_{tk}} f(x)}{|S_{tk}|} \tag{2}$$

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

$$\mathbb{E}[(\hat{\mu}_t - \mu_t)^2] = \mathbb{E}[\hat{\mu}_t - \mu_t]^2 + Var[\hat{\mu}_t] \tag{3}$$

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:

$$\mu_t = \mathbb{E}_{x \sim \mathcal{D}_t^+}[f(x)|x \in \mathcal{D}_t^+] = \sum_{x \in \mathcal{D}_t^+} \frac{1}{|\mathcal{D}_t^+|} \cdot f(x) \tag{4}$$

$$= \frac{1}{|\mathcal{D}_t^+|} \sum_{x \in \mathcal{D}_t^+} f(x) \tag{5}$$

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

$$\mu_t = \frac{1}{|\mathcal{D}_t^+|}\left(\sum_{x\in\mathcal{D}_{t1}^+} f(x) + \cdots + \sum_{x\in\mathcal{D}_{tK}^+} f(x)\right) \tag{6}$$

$$= \frac{1}{|\mathcal{D}_t^+|}\left(\frac{|\mathcal{D}_{t1}^+|}{|\mathcal{D}_{t1}^+|}\sum_{x\in\mathcal{D}_{t1}^+} f(x) + \cdots + \frac{|\mathcal{D}_{tK}^+|}{|\mathcal{D}_{tK}^+|}\sum_{x\in\mathcal{D}_{tK}^+} f(x)\right) \tag{7}$$

$$= \frac{1}{|\mathcal{D}_t^+|}\Big(|\mathcal{D}_{t1}^+|\cdot\mu_{t1} + \cdots + |\mathcal{D}_{tK}^+|\cdot\mu_{tK}\Big) \tag{8}$$

$$= \frac{1}{|\mathcal{D}_t^+|}\sum_{k=1}^{K}|\mathcal{D}_{tk}^+|\cdot\mu_{tk} \tag{9}$$

$$\mu_t = \sum_{k=1}^{K}\frac{|\mathcal{D}_{tk}^+|}{|\mathcal{D}_t^+|}\cdot\mu_{tk} \tag{10}$$

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

$$\mu_t = \sum_{k=1}^{K}\frac{|\mathcal{D}_{tk}|p_{tk}}{\sum_{j=1}^K |\mathcal{D}_{tj}|p_{tj}}\cdot\mu_{tk} \tag{11}$$

$$\mu_t = \sum_{k=1}^{K} w_{tk}\cdot\mu_{tk} \tag{12}$$

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:

$$\mathbb{E}_{x\sim\mathcal{D}_t^+}[\hat{\mu}_t] = \mathbb{E}_{x\sim\mathcal{D}_t^+}\left[\sum_{k=1}^{K} w_{tk}\cdot\frac{\sum_{x\in S_{tk}} f(x)}{|S_{tk}|}\right] \tag{13}$$

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

$$\mathbb{E}_{x\sim\mathcal{D}_t}[\hat{\mu}_t] = \sum_{k=1}^{K} w_{tk}\cdot\mathbb{E}_{x\sim S_{tk}}\left[\frac{\sum_{x\in S_{tk}} f(x)}{|S_{tk}|}\right] \tag{14}$$

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

$$\mathbb{E}_{x\sim\mathcal{D}_t}[\hat{\mu}_t] = \sum_{k=1}^{K} w_{tk}\cdot\frac{1}{|S_{tk}|}\cdot\sum_{x\in S_{tk}}\mathbb{E}_{x\sim\mathcal{S}_{tk}}[f(x)] \tag{15}$$

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

$$\mathbb{E}_{x \sim \mathcal{D}_t}[\hat{\mu}_t] = \sum_{k=1}^{K} w_{tk} \cdot \frac{1}{|S_{tk}|} \cdot \sum_{x \in S_{tk}} \mu_{tk} \tag{16}$$

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

$$= \sum_{k=1}^{K} w_{tk} \cdot \mu_{tk} \tag{18}$$

Looking back at equation (11), 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 $Var[\hat{\mu}_t] = \sum_{k=1}^{K} \frac{w_{tk}^2 \sigma_{tk}^2}{p_{tk}\left(\frac{N_1}{K} + N_2 a_{tk}\right)}$.

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

$$Var[\hat{\mu}_t] = Var\left[\sum_{k=1}^{K} w_{tk} \cdot \frac{\sum_{x \in S_{tk}} f(x)}{|S_{tk}|}\right] \tag{19}$$

$$= \sum_{k=1}^{K} Var\left[w_{tk} \cdot \frac{\sum_{x \in S_{tk}} f(x)}{|S_{tk}|}\right] \tag{20}$$

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

$$Var[\hat{\mu}_t] = \sum_{k=1}^{K} \left(\frac{w_{tk}}{|S_{tk}|}\right)^2 \cdot \sum_{x \in S_{tk}} Var[f(x)] \tag{21}$$

Each sample $x \in S_{tk}$ 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:

$$Var[\hat{\mu}_t] = \sum_{k=1}^{K} \left(\frac{w_{tk}}{|S_{tk}|}\right)^2 \cdot \sum_{x \in S_{tk}} \sigma_{tk}^2 \tag{22}$$

$$= \sum_{k=1}^{K} \left(\frac{w_{tk}}{|S_{tk}|}\right)^2 \cdot |S_{tk}| \cdot \sigma_{tk}^2 \tag{23}$$

$$= \sum_{k=1}^{K} \frac{w_{tk}^2 \sigma_{tk}^2}{|S_{tk}|} \tag{24}$$

Finally, plugging in $|S_{tk}| = p_{tk}\big(\frac{N_1}{K} + N_2 a_{tk}\big)$ gives us our expected error for $\hat{\mu}_t$:

$$Var[\hat{\mu}_t] = \sum_{k=1}^{K} \frac{w_{tk}^2 \sigma_{tk}^2}{p_{tk}\big(\frac{N_1}{K} + N_2 a_{tk}\big)} \tag{25}$$

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:

$$\mathbb{E}[(\hat{\mu}_t - \mu_t)^2] = Var[\hat{\mu}_t] \tag{26}$$

$$\mathbb{E}[(\hat{\mu}_t - \mu_t)^2] = \sum_{k=1}^{K} \frac{w_{tk}^2 \sigma_{tk}^2}{p_{tk}\big(\frac{N_1}{K} + N_2 a_{tk}\big)} \tag{27}$$

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 $|S_{tk}| = p_{tk}\big(\frac{N_1}{K} + N_2 a_{tk}\big)$.

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

$$a_{tk}^* = \frac{|\mathcal{D}_{tk}|\sqrt{p_{tk}}\sigma_{tk}}{\frac{N_2}{N}\sum_{j=1}^{K}|\mathcal{D}_{tj}|\sqrt{p_{tj}}\sigma_{tj}} - \frac{N_1}{N_2 K} \tag{28}$$

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:

$$\mathcal{L}(a_{t1}, \ldots, a_{tK}) = \sum_{k=1}^{K} \frac{w_{tk}^2 \sigma_{tk}^2}{p_{tk}\big(\frac{N_1}{K} + N_2 a_{tk}\big)} \tag{29}$$

$$g(a_{t1}, \ldots, a_{tK}) = \sum_{k=1}^{K} a_{tk} - 1 \tag{30}$$

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

$$\frac{\partial \mathcal{L}}{\partial a_{t1}} = \lambda \frac{\partial g}{\partial a_{t1}} \tag{31}$$

$$\dots \tag{32}$$

$$\frac{\partial \mathcal{L}}{\partial a_{tK}} = \lambda \frac{\partial g}{\partial a_{tK}} \tag{33}$$

$$g(a_{t1}, \dots, a_{tK}) = 1 \tag{34}$$

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:

$$\frac{\partial}{\partial a_{ti}} \left(\sum_{k=1}^{K} \frac{w_{tk}^2 \sigma_{tk}^2}{p_{tk}\left(\frac{N_1}{K} + N_2 a_{tk}\right)} \right) = \lambda \frac{\partial}{\partial a_{ti}} \left(\sum_{k=1}^{K} a_{tk} - 1 \right) \tag{35}$$

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

$$-\frac{w_{ti}^2 \sigma_{ti}^2 N_2}{p_{ti}\left(\frac{N_1}{K} + N_2 a_{ti}\right)^2} = \lambda \tag{36}$$

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

$$\frac{w_{ti}^2 \sigma_{ti}^2 N_2}{p_{ti}\left(\frac{N_1}{K} + N_2 a_{ti}\right)^2} = \lambda \tag{37}$$

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

$$a_{ti} = \frac{w_{ti} \sigma_{ti}}{\sqrt{\lambda N_2 p_{ti}}} - \frac{N_1}{N_2 K} \tag{38}$$

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

$$\sum_{k=1}^{K} a_{tk} = 1 \tag{39}$$

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

Solving for λ gives us:

$$\lambda = \left(\sum_{k=1}^{K} \frac{w_{tk} \sigma_{tk} \sqrt{N_2}}{\sqrt{p_{tk}} N} \right)^2 \tag{41}$$

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}^*:

$$a_{tk}^* = \frac{w_{tk}\sigma_{tk}}{\sqrt{N_2 p_{tk}}\left(\sum_{j=1}^{K} \frac{w_{tj}\sigma_{tj}\sqrt{N_2}}{\sqrt{p_{tj}}N}\right)} - \frac{N_1}{N_2 K} \tag{42}$$

$$a_{tk}^* = \frac{w_{tk}\sigma_{tk}}{\frac{N_2}{N}\sqrt{p_{tk}}\sum_{j=1}^{K} \frac{w_{tj}\sigma_{tj}}{\sqrt{p_{tj}}}} - \frac{N_1}{N_2 K} \tag{43}$$

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

$$a_{tk}^* = \frac{|\mathcal{D}_{tk}| p_{tk}\sigma_{tk}}{\frac{N_2}{N}\sqrt{p_{tk}}\sum_{j=1}^{K} \frac{\mathcal{D}_{tj}|p_{tj}\sigma_{tj}}{\sqrt{p_{tj}}}} - \frac{N_1}{N_2 K} \tag{44}$$

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

$$a_{tk}^* = \frac{|\mathcal{D}_{tk}|\sqrt{p_{tk}}\sigma_{tk}}{\frac{N_2}{N}\sum_{j=1}^{K} |\mathcal{D}_{tj}|\sqrt{p_{tj}}\sigma_{tj}} - \frac{N_1}{N_2 K} \tag{45}$$

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 Expected Error under 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:

$$\hat{a}_{tk} = \frac{|\mathcal{D}_{<tk}|\sqrt{\hat{p}_{<tk}}\hat{\sigma}_{<tk}}{\frac{N_2}{N}\sum_{j=1}^{K} |\mathcal{D}_{<tj}|\sqrt{\hat{p}_{<tj}}\hat{\sigma}_{<tj}} - \frac{N_1}{N_2 K} \tag{46}$$

Theorem 3. Given the assumptions stated above, in the limit as $t \to \infty$:

$$\mathbb{E}[(\hat{\mu}_t - \mu_t)^2] \to \mathbb{E}[(\hat{\mu}_t^* - \mu_t)^2] \tag{47}$$

Proof. We define $\mathbb{E}_{x\sim\mathcal{D}_t^+}[f(x)] = \mu_t$. We wish to show that the error $\mathbb{E}[(\hat{\mu}_t - \mu_t)^2]$ of our estimator $\hat{\mu}_t$ approaches the optimal error $\mathbb{E}[(\hat{\mu}_t^* - \mu_t)^2]$ in the limit as t goes to infinity under the assumptions stated above. Our estimator and the optimal estimator are defined as follows:

$$\hat{\mu}_t = \sum_{k=1}^{K} \hat{w}_{tk} \cdot \frac{\sum_{x\in S_{tk}} f(x)}{\hat{p}_{tk}\big(\frac{N_1}{K} + \hat{a}_{tk}N_2\big)} \tag{48}$$

$$\hat{\mu}_t^* = \sum_{k=1}^{K} \hat{w}_{tk} \cdot \frac{\sum_{x\in S_{tk}} f(x)}{\hat{p}_{tk}\big(\frac{N_1}{K} + a_{tk}^*N_2\big)} \tag{49}$$

Due to the symmetry of the equations above, in order to show that:

$$\mathbb{E}[(\hat{\mu}_t - \mu_t)^2] \to \mathbb{E}[(\hat{\mu}_t^* - \mu_t)^2] \tag{50}$$

In the limit as $t \to \infty$, it suffices to show that:

$$\hat{a}_{tk} \to a_{tk}^* \tag{51}$$

From equation (46), we can see that $\hat{a}_{tk}$ is a function of two random variables: $\hat{p}_{<tk}$ and $\hat{\sigma}_{<tk}$. Thus, in order to show that $\hat{a}_{tk} \to a_{tk}^*$ in the limit as $t \to \infty$, we'll begin our proof by putting concentration inequalities on $\hat{p}_{<tk}$ and $\hat{\sigma}_{<tk}$ in four separate lemmas. We'll then show that the concentration inequalities force $\hat{p}_{<tk} \to p_{tk}$ and $\hat{\sigma}_{<tk} \to \sigma_{tk}$ as t goes to infinity.

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,*

$$\sqrt{\hat{p}_{<tk}} \le \sqrt{p_{<tk}} \cdot \sqrt{1 + O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)} \tag{52}$$

Proof. Let us first define $N_{<t} = (t-1)N_1/K$. We apply Chung and Lu [2002] 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,

$$P(\hat{p}_{tk} \ge p_{tk} + \lambda) \le exp\left(\frac{-\lambda^2}{2(\frac{p_{<tk}}{N_{<t}} + \frac{\lambda}{3N_{<t}})}\right) \tag{53}$$

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 $1-\delta$,

$$\hat{p}_{<tk} \leq p_{<tk} + \frac{\sqrt{18N_{<t}p_{<tk}\ln(1/\delta) + \ln(1/\delta)^2} - \ln(1/\delta)}{3N_{<t}} \tag{54}$$

$$\leq p_{<tk} + \frac{\sqrt{18N_{<t}p_{<tk}\ln(1/\delta)}}{3N_{<t}} = p_{<tk} + \sqrt{\frac{2\ln(1/\delta)p_{<tk}}{N_{<t}}} \tag{55}$$

With probability at least $1 - \delta$,

$$\sqrt{\hat{p}_{<tk}} \leq \sqrt{p_{<tk} + \sqrt{\frac{2\ln(1/\delta)p_{<tk}}{N_{<t}}}} \leq \sqrt{p_{<tk}} \cdot \sqrt{1 + \sqrt{\frac{2\ln(1/\delta)}{p_{<tk}N_{<t}}}} \tag{56}$$

$$\sqrt{\hat{p}_{<tk}} \leq \sqrt{p_{<tk}} \cdot \sqrt{1 + O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)} \tag{57}$$

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,*

$$\sqrt{\hat{p}_{<tk}} \geq \sqrt{p_{<tk}} \cdot \sqrt{1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)} \tag{58}$$

Proof. Let us first define $N_{<t} = (t-1)N_1/K$. We apply Chung and Lu [2002] 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,

$$P(\hat{p}_{tk} < p_{tk} - \lambda) \leq exp\left(\frac{-\lambda^2}{2(\frac{p_{<tk}}{N_{<t}})}\right) \tag{59}$$

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 $1 - \delta$,

$$\hat{p}_{<tk} \geq p_{<tk} - \frac{\sqrt{p_{<tk}}}{\sqrt{N_{<t}}} \cdot \sqrt{2\ln(1/\delta)} = p_{<tk} - \sqrt{\frac{2\ln(1/\delta)p_{<tk}}{N_{<t}}} \tag{60}$$

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

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

$$\geq \sqrt{p_{<tk}}\sqrt{1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)} \tag{62}$$

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 $|S_{<tk}|$.

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

$$|S_{<tk}| \geq p_{<tk}N_{<t} - \sqrt{2\ln(1/\delta)p_{<tk}N_{<t}} \tag{63}$$

Proof. Let us first define $N_{<t} = (t-1)N_1/K$. We apply Tarjan [2009] reproduced as Lemma 7.2 to lower bound $|S_{<tk}|$. Note that $|S_{<tk}| \sim Binomial(N_{<t}, p_{<tk})$ and that equivalently $|S_{<tk}| = \sum_{i=1}^{N_{<t}} Y_i$ where $Y_i \sim Bernoulli(p_{<tk})$. Hence,

$$P\Big(|S_{<tk}| \leq (1-\epsilon)p_{<tk}N_{<t}\Big) \leq exp\Big(\frac{-\epsilon^2 p_{<tk}N_{<t}}{2}\Big) \tag{64}$$

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 $1-\delta$:

$$|S_{<tk}| \geq p_{<tk}N_{<t} - \sqrt{2\ln(1/\delta)p_{<tk}N_{<t}} \tag{65}$$

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 $|S_{<tk}| \geq 2$). *For small $\delta > 0$, the following holds for all k simultaneously,*

$$\sigma_{<tk}\sqrt{1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)} \leq \hat{\sigma}_{<tk} \leq \sigma_{<tk}\sqrt{1 + O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)} \tag{66}$$

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 $|S_{<tk}| \geq 2$, which is:

$$\hat{\sigma}^2_{tk} = \frac{1}{|S_{<tk}| - 1} \sum_{i=1}^{|S_{<tk}|} (X_{tk,i} - \hat{\mu}_{tk})^2 \tag{67}$$

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 [2005] reproduced as Lemma 7.3). For bounded U-statistics, we can apply bounded differences to obtain a concentration bound (Rinaldo [2018] reproduced as Lemma 7.4). Namely, where $g(X_{tk,i}, X_{tk,j}) \leq b$,

$$P(|U_n - \mathbb{E}[U_n]| \geq t) \leq 2\,\exp\Big(-\frac{nt^2}{8b^2}\Big) \tag{68}$$

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

$$P(|\hat{\sigma}_{<tk} - \sigma_{<tk}| \geq t) \leq 2\,\exp\Bigg(-\frac{|S_{<tk}|t^2}{8C_k^{(\mu_t^4)}}\Bigg) \tag{69}$$

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

$$\sigma^2_{<tk} - \sqrt{\frac{8\ln(1/\delta)C_k^{(\mu_t^4)}}{|S_{<tk}|}} \leq \hat{\sigma}^2_{<tk} \leq \sigma^2_{<tk} + \sqrt{\frac{8\ln(1/\delta)C_k^{(\mu_t^4)}}{|S_{<tk}|}} \tag{70}$$

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

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

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

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

$$\sigma^2_{<tk} - \sqrt{\frac{8\ln(1/\delta)C_k^{(\mu_t^4)}}{p_{<tk}N_{<t}} + \frac{8\sqrt{2}\ln(1/\delta)^{3/2}}{p_{<tk}^{3/2}N_{<t}^{3/2}}} \leq \hat{\sigma}^2_{<tk} \leq \sigma^2_{<tk} + \sqrt{\frac{8\ln(1/\delta)C_k^{(\mu_t^4)}}{p_{<tk}N_{<t}} + \frac{8\sqrt{2}\ln(1/\delta)^{3/2}}{p_{<tk}^{3/2}N_{<t}^{3/2}}} \tag{74}$$

$$\sigma^2_{<tk} - \sqrt{O\left(\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}\right)} \leq \hat{\sigma}^2_{<tk} \leq \sigma^2_{<tk} + \sqrt{O\left(\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}\right)} \tag{75}$$

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

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

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

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 (). However, if $\sigma_{<tk} = 0$, then $\hat{\sigma}_{<tk} = \sigma_{<tk}$, so the bound still holds.

5.0.3 Final Result

Recall that $N_{<t} = (t-1)N_1/K$. In the limit as $t \to \infty$, we can see that the big-O terms on the left and right-hand sides of the concentration inequalities

both go to 0 due to the factor of t in their denominators. As a result, we have that:

$$\lim_{t\to\infty}\hat{p}_{<tk} = p_{<tk} \tag{78}$$
$$\lim_{t\to\infty}\hat{\sigma}_{<tk} = \sigma_{<tk} \tag{79}$$

Therefore, plugging these into our expression for $\hat{a}_{tk}$ we have that:

$$\lim_{t\to\infty}\hat{a}_{tk} = \frac{|\mathcal{D}_{<tk}|\sqrt{p_{<tk}}\sigma_{<tk}}{\frac{N_2}{N}\sum_{j=1}^{K}|\mathcal{D}_{<tj}|\sqrt{p_{<tj}}\sigma_{<tj}} - \frac{N_1}{N_2 K} \tag{80}$$

Finally, because we assumed that our dataset follows a stationary distribution, we have that $|\mathcal{D}_{<tk}| = |\mathcal{D}_{tk}|$, $p_{<tk} = p_{tk}$, and $\sigma_{<tk} = \sigma_{tk}$, therefore:

$$\lim_{t\to\infty}\hat{a}_{tk} = \frac{|\mathcal{D}_{tk}|\sqrt{p_{tk}}\sigma_{tk}}{\frac{N_2}{N}\sum_{j=1}^{K}|\mathcal{D}_{tj}|\sqrt{p_{tj}}\sigma_{tj}} - \frac{N_1}{N_2 K} \tag{81}$$

This is equivalent to our expression for a^*_{tk} in equation 1. This concludes our proof that our estimator's error approaches the optimal error in the limit as t goes to infinity.

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:

$$\mathbb{E}[(\hat{\mu}_t - \mu_t)^2] \tag{82}$$

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:

$$\hat{a}_{tk} = \frac{|\mathcal{D}_{<tk}|\sqrt{\hat{p}_{<tk}}\hat{\sigma}_{<tk}}{\frac{N_2}{N}\sum_{j=1}^{K}|\mathcal{D}_{<tj}|\sqrt{\hat{p}_{<tj}}\hat{\sigma}_{<tj}} - \frac{N_1}{N_2 K} \tag{83}$$

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

$$\mathbb{E}[(\hat{\mu}_t - \mu_t)^2] \leq O\left(\frac{1}{N_1}\right) + O\left(\frac{N_1}{N_2^2}\right) + O\left(\frac{1}{N_2\sqrt{N_1}}\right) \tag{84}$$

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,*

$$\sqrt{\hat{p}_{tk}} \leq \sqrt{p_{tk}} \cdot \sqrt{1 + O\left(\sqrt{\frac{\ln(1/\delta)}{p_{tk}N_1}}\right)} \tag{85}$$

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,*

$$\sqrt{\hat{p}_{tk}} \geq \sqrt{p_{tk}} \cdot \sqrt{1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{tk}N_1}}\right)} \tag{86}$$

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,*

$$w_{tk} - O\left(\frac{\ln(1/\delta)\sqrt{p_{tk}}}{\sqrt{N_1}}\right) \leq \hat{w}_{tk} \leq w_{tk} + O\left(\frac{\ln(1/\delta)\sqrt{p_{tk}}}{\sqrt{N_1}}\right) \tag{87}$$

Proof. Recall that $\hat{w}_{tk} = |\mathcal{D}_{tk}|p_{tk}/\sum_{j=1}^{K}|\mathcal{D}_{tj}|p_{tj}$. Equation (87) 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 (87). Specifically, the condition is that for each k,

$$p_{tk} - \sqrt{\frac{2\ln(1/\delta)p_{tk}}{N_1}} \leq \hat{p}_{tk} \leq p_{tk} + \sqrt{\frac{2\ln(1/\delta)p_{tk}}{N_1}} \tag{88}$$

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

$$\frac{|\mathcal{D}_{tk}|\big(p_{tk} - \sqrt{\frac{2\ln(1/\delta)p_{tk}}{N_1}}\big)}{\sum_{j=1}^{K} |\mathcal{D}_{tj}|\big(p_{tj} + \sqrt{\frac{2\ln(1/\delta)p_{tj}}{N_1}}\big)} \leq \hat{w}_{tk} \leq \frac{|\mathcal{D}_{tk}|\big(p_{tk} + \sqrt{\frac{2\ln(1/\delta)p_{tk}}{N_1}}\big)}{\sum_{j=1}^{K} |\mathcal{D}_{tj}|\big(p_{tj} - \sqrt{\frac{2\ln(1/\delta)p_{tj}}{N_1}}\big)} \tag{89}$$

We now simplify this bound with Big-O notation.

$$\begin{aligned}
\hat{w}_{tk} &\leq \frac{|\mathcal{D}_{tk}|\big(p_{tk} + \sqrt{\frac{2\ln(1/\delta)p_{tk}}{N_1}}\big)}{\sum_{j=1}^{K} |\mathcal{D}_{tj}|\big(p_{tj} - \sqrt{\frac{2\ln(1/\delta)p_{tj}}{N_1}}\big)} && (90)\\
&\leq \frac{|\mathcal{D}_{tk}|p_{tk}\big(1 + \sqrt{\frac{2\ln(1/\delta)}{p_{tk}N_1}}\big)}{\sum_{j=1}^{K} |\mathcal{D}_{tj}|p_{tj}\big(1 - \sqrt{\frac{2\ln(1/\delta)}{p_{tj}N_1}}\big)} && (91)\\
&\leq \frac{\frac{1}{\sum_{j=1}^{K}|\mathcal{D}_{tj}|p_{tj}} \cdot |\mathcal{D}_{tk}|p_{tk}\big(1 + \sqrt{\frac{2\ln(1/\delta)}{p_{tk}N_1}}\big)}{\frac{1}{\sum_{j=1}^{K}|\mathcal{D}_{tj}|p_{tj}} \cdot \sum_{j=1}^{K} |\mathcal{D}_{tj}|p_{tj}\big(1 - \sqrt{\frac{2\ln(1/\delta)}{p_{tj}N_1}}\big)} && (92)\\
&\leq \frac{w_{tk}\left(1 + \sqrt{\frac{2\ln(1/\delta)}{p_{tk}N_1}}\right)}{1 - \sqrt{\frac{2\ln(1/\delta)}{p_{tk}N_1}}} && (93)\\
&\leq w_{tk}\left(1 + \sqrt{\frac{2\ln(1/\delta)}{p_{tk}N_1}}\right)^2 && (94)\\
&\leq w_{tk}\left(1 + 2\sqrt{\frac{2\ln(1/\delta)}{p_{tk}N_1}} + \frac{2\ln(1/\delta)}{p_{tk}N_1}\right) && (95)\\
&\leq w_{tk} + O\left(\sqrt{\frac{\ln(1/\delta)p_{tk}}{N_1}}\right) + O\left(\frac{\ln(1/\delta)}{N_1}\right) && (96)\\
&\leq w_{tk} + O\left(\frac{\ln(1/\delta)\sqrt{p_{tk}}}{\sqrt{N_1}}\right) && (97)
\end{aligned}$$

We repeat these steps for the lower bound. Thus:

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

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

$$\sqrt{\hat{p}_{<tk}}\hat{\sigma}_{<tk} \leq \sqrt{p_{<tk}}\sigma_{<tk} + O\left(\frac{1}{\sqrt{N_{<t}}}\right) \tag{99}$$

Proof. In the cases where $|S_{<tk}| < 2$, 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.

$$\sqrt{\hat{p}_{<tk}}\hat{\sigma}_{<tk} \leq \sqrt{\frac{2}{N_{<t}}}\hat{\sigma}_{<tk} \leq \sqrt{p_{<tk}}\sigma_{<tk} + \sqrt{\frac{2}{N_{<t}}}C_k^{(\sigma)} \tag{100}$$

$$\sqrt{\hat{p}_{<tk}}\hat{\sigma}_{<tk} \leq \sqrt{p_{<tk}}\sigma_{<tk} + O\left(\frac{1}{\sqrt{N_{<t}}}\right) \tag{101}$$

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

$$\hat{a}_{tk} \geq a_{tk}\left(1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{tk}(t-1)N_1}}\right)\right) \tag{102}$$

Proof. Our goal is to lower bound:

$$\hat{a}_{tk} = \frac{|\mathcal{D}_{<tk}|\sqrt{\hat{p}_{<tk}}\hat{\sigma}_{<tk}}{\frac{N_2}{N}\sum_{j=1}^{K}|\mathcal{D}_{<tj}|\sqrt{\hat{p}_{<tj}}\hat{\sigma}_{<tj}} - \frac{N_1}{N_2 K} \tag{103}$$

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 $|S_{<tk}| \geq 2$. Through Lemma 5.3, we know that for $|S_{<tk}| > 2$ to be true, we can condition on $p_{<tk} > p^*_< = \frac{2\ln(1/\delta)+2\sqrt{\ln(1/\delta)}+2}{|S_{<tk}|}$. This allows us to satisfy the conditions of Lemma 5.3.

$$\sqrt{\hat{p}_{<tk}}\hat{\sigma}_{<tk} \geq \sqrt{p_{<tk}}\sigma_{<tk}\sqrt{1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)}\sqrt{1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)} \tag{104}$$

$$\geq \sqrt{p_{<tk}}\sigma_{<tk}\left(1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)\right) \tag{105}$$

$$\geq \sqrt{p_{<tk}}\sigma_{<tk} - O\left(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\right) \tag{106}$$

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

$$\sum_{k=1}^{K}\sqrt{\hat{p}_{<tk}}\hat{\sigma}_{<tk} \leq \sum_{p_{<tk}>p^*_<}^{K}\sqrt{p_{<tk}}\sigma_{<tk}\left(1+O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)\right)+\sum_{p_{<tk}\leq p^*_<}\sqrt{p_{<tk}}\sigma_{<tk}+O\left(\frac{1}{\sqrt{N_{<t}}}\right) \tag{107}$$

$$\leq \sum_{k=1}^{K}\sqrt{p_{<tk}}\sigma_{<tk}+O\left(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\right) \tag{108}$$

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

$$\hat{a}_{tk}+\frac{N_1}{N_2K} \geq \frac{|\mathcal{D}_{<tk}|\sqrt{p_{<tk}}\sigma_{<tk}-O\left(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\right)}{\frac{N_2}{N}\sum_{j=1}^{K}|\mathcal{D}_{<tj}|\sqrt{p_{<tj}}\sigma_{<tj}+O\left(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\right)} \tag{109}$$

$$\geq \frac{(a_{tk}+\frac{N_1}{N_2K})-O\left(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\right)}{1+O\left(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\right)} \tag{110}$$

$$\geq \left((a_{tk}+\frac{N_1}{N_2K})-O\left(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\right)\right)\left(1-O\left(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\right)\right) \tag{111}$$

$$\geq \left(a_{tk}+\frac{N_1}{N_2K}\right)-O\left(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\right)-O\left(\sqrt{\frac{p_{<tk}\ln(1/\delta)}{N_{<t}}}\right)+O\left(\frac{\ln(1/\delta)}{N_{<t}}\right) \tag{112}$$

$$\geq \left(a_{tk}+\frac{N_1}{N_2K}\right)-O\left(\sqrt{\frac{\ln(1/\delta)}{N_{<t}}}\right) \tag{113}$$

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

$$\hat{a}_{tk} \geq a_{tk}-O\left(\sqrt{\frac{\ln(1/\delta)}{(t-1)N_1}}\right) \tag{114}$$

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

$$\hat{a}_{tk} \geq a_{tk}\left(1-O\left(\sqrt{\frac{\ln(1/\delta)}{p_{tk}(t-1)N_1}}\right)\right) \tag{115}$$

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

$$|S_{tk}| \geq p_{tk}a_{tk}N_2\left(1 - O\left(\sqrt{\frac{1}{p_{tk}(t-1)N_1}}\right) - O\left(\sqrt{\frac{\ln(1/\gamma)}{p_{tk}^{3/2}N_2}}\right)\right) \tag{116}$$

Proof. Recall that $|S_{tk}| \sim Binomial(\frac{N_1}{K} + \hat{a}_{tk}N_2)$. For simplicity, let us ignore the defensive samples – which will not invalidate the lower bound we derive – and instead use $|S_{tk}| \sim Binomial(\hat{a}_{tk}N_2)$. Thus, using Tarjan [2009] reproduced as Lemma 7.2 we have that:

$$P\Big(|S_{tk}| \leq (1-\epsilon)p_{tk}\hat{a}_{tk}N_2\Big) \leq exp\left(\frac{-\epsilon^2 p_{tk}\hat{a}_{tk}N_2}{2}\right) \tag{117}$$

We set $\gamma = exp\left(\frac{-\epsilon^2 p_{tk}\hat{a}_{tk}N_2}{2}\right)$ 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:

$$|S_{tk}| \geq p_{tk}\hat{a}_{tk}N_2\left(1 - \sqrt{\frac{2\ln(1/\gamma)}{p_{tk}\hat{a}_{tk}N_2}}\right) \tag{118}$$

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

$$\geq p_{tk}a_{tk}N_2\left(1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{tk}N_{<t}}}\right)\right)\left(1 - \sqrt{\frac{2\ln(1/\gamma)}{p_{tk}a_{tk}N_2\left(1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{tk}N_{<t}}}\right)\right)}}\right) \tag{119}$$

$$\geq p_{tk}a_{tk}N_2\left(1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{tk}N_{<t}}}\right) - \sqrt{\frac{2\ln(1/\gamma)}{p_{tk}a_{tk}N_2\left(1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{tk}N_{<t}}}\right)\right)}}\right) \tag{120}$$

$$\geq p_{tk}a_{tk}N_2\left(1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{tk}N_{<t}}}\right) - \sqrt{\frac{2\ln(1/\gamma)}{p_{tk}a_{tk}N_2} \cdot \left(1 + O\left(\sqrt{\frac{\ln(1/\delta)}{p_{tk}N_{<t}}}\right)\right)}\right) \tag{121}$$

$$\geq p_{tk}a_{tk}N_2\left(1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{tk}N_{<t}}}\right) - \sqrt{\frac{2\ln(1/\gamma)}{p_{tk}a_{tk}N_2} + O\left(\sqrt{\frac{\ln(1/\delta)\ln(1/\gamma)}{p_{tk}^3 N_2^2 p_{tk}N_{<t}}}\right)}\right) \tag{122}$$

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

$$|S_{tk}| \geq p_{tk} a_{tk} N_2 \left(1 - O\left(\sqrt{\frac{1}{p_{tk}(t-1)N_1}}\right) - O\left(\sqrt{\frac{\ln(1/\gamma)}{p_{tk}^{3/2} N_2}}\right)\right) \tag{123}$$

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

$$p_{tk}^x P(|S_{tk}|=0) \leq O\left(\frac{1}{N_1^x}\right) + O\left(\frac{\sqrt{N_1}}{N_2^2}\right) \tag{124}$$

Proof.

$$p_{tk}^x P(|S_{tk}|=0) = p_{tk}^x (1-p_{tk})^{\lceil \frac{N_1}{K} + \hat{a}_{tk} N_2 \rceil} \leq p_{tk}^x (1-p_{tk})^{(\frac{N_1}{K} + \hat{a}_{tk} N_2)} \tag{125}$$

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

$$p_{tk}^x (1-p_{tk})^{(\frac{N_1}{K} + \hat{a}_{tk} N_2)} \leq O\left(\frac{1}{N_1^x}\right) \tag{126}$$

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

$$p_{tk}^x (1-p_{tk})^{(\frac{N_1}{K} + \hat{a}_{tk} N_2)} \leq (1-p_{tk})^{\left(\frac{N_1}{K} + a_{tk} N_2 \left(1 - O\left(\frac{1}{\sqrt{p_{tk}(t-1)N_1}}\right)\right)\right)} \tag{127}$$

$$\leq (1-p_{tk})^{\left(\frac{N_1}{K} + N_2 \left(O(\sqrt{p_{tk}}) - O\left(\frac{1}{\sqrt{(t-1)N_1}}\right)\right)\right)} \tag{128}$$

$$\tag{129}$$

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

$$p_{tk}^x (1-p_{tk})^{(\frac{N_1}{K} + \hat{a}_{tk} N_2)} \leq (1-p_{tk})^{\left(\frac{N_1}{K} + N_2 \left(O(\frac{1}{\sqrt{N_1}}) - O\left(\frac{1}{\sqrt{(t-1)N_1}}\right)\right)\right)} \tag{130}$$

$$\leq (1-p_{tk})^{\left(\frac{N_1}{K} + O\left(\frac{N_2}{N_1^{1/2}}\right)\right)} \tag{131}$$

$$\leq (1-p_{tk})^{O\left(\frac{N_2}{N_1^{1/2}}\right)} \tag{132}$$

We assume that $N_2 = w(N_1^{1/2})$ which means that (127) is decreasing exponentially. Hence we can say that $(1-p_{tk})^{O\left(\frac{N_2}{N_1^{1/2}}\right)} = O\left(\frac{N_1}{N_2^2}\right)$:

$$p_{tk}^x P(|S_{tk}|=0) \leq O\left(\frac{1}{N_1^x}\right) + O\left(\frac{N_1}{N_2^2}\right) \tag{133}$$

6.2 Final Result

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

$$\sqrt{\hat{p}_{<tk}} \leq \sqrt{p_{<tk}} \cdot \sqrt{1 + O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)} \tag{134}$$

$$\sqrt{\hat{p}_{<tk}} \geq \sqrt{p_{<tk}} \cdot \sqrt{1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)} \tag{135}$$

$$|S_{<tk}| \geq p_{<tk}N_{<t} - \sqrt{2\ln(1/\delta)p_{<tk}N_{<t}} \tag{136}$$

$$\sigma_{<tk}\sqrt{1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)} \leq \hat{\sigma}_{<tk} \leq \sigma_{<tk}\sqrt{1 + O\left(\sqrt{\frac{\ln(1/\delta)}{p_{<tk}N_{<t}}}\right)} \tag{137}$$

$$\sqrt{\hat{p}_{tk}} \leq \sqrt{p_{tk}} \cdot \sqrt{1 + O\left(\sqrt{\frac{\ln(1/\delta)}{p_{tk}N_1}}\right)} \tag{138}$$

$$\sqrt{\hat{p}_{tk}} \geq \sqrt{p_{tk}} \cdot \sqrt{1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{tk}N_1}}\right)} \tag{139}$$

$$w_{tk} - O\left(\frac{\ln(1/\delta)\sqrt{p_{tk}}}{\sqrt{N_1}}\right) \leq \hat{w}_{tk} \leq w_{tk} + O\left(\frac{\ln(1/\delta)\sqrt{p_{tk}}}{\sqrt{N_1}}\right) \tag{140}$$

$$\sqrt{\hat{p}_{<tk}}\hat{\sigma}_{<tk} \leq \sqrt{p_{<tk}}\sigma_{<tk} + O\left(\frac{1}{\sqrt{N_{<t}}}\right) \tag{141}$$

$$\hat{a}_{tk} \geq a_{tk}\left(1 - O\left(\sqrt{\frac{\ln(1/\delta)}{p_{tk}(t-1)N_1}}\right)\right) \tag{142}$$

$$|S_{tk}| \geq p_{tk}a_{tk}N_2\left(1 - O\left(\sqrt{\frac{1}{p_{tk}(t-1)N_1}}\right) - O\left(\sqrt{\frac{\ln(1/\gamma)}{p_{tk}^{3/2}N_2}}\right)\right) \tag{143}$$

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

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:

$$\mathbb{E}[(\hat{\mu}_t - \mu_t)^2] = \mathbb{E}\Big[\Big(\sum_{k=1}^{K} \hat{w}_{tk}\hat{\mu}_{tk} - w_{tk}\mu_{tk}\Big)^2\Big] \tag{145}$$

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

$$= \mathbb{E}\Big[\Big(\sum_{k=1}^{K} \hat{w}_{tk}(\hat{\mu}_{tk} - \mathbb{E}[\hat{\mu}_{tk}]) + (\hat{w}_{tk} - w_{tk})\mathbb{E}[\hat{\mu}_{tk}] - w_{tk}\mu_{tk}P(|S_{tk}| = 0)\Big)^2\Big] \tag{146}$$

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 $B = (\hat{w}_{tk} - w_{tk})\mathbb{E}[\hat{\mu}_{tk}] - w_{tk}\mu_{tk}P(|S_{tk}| = 0)$. Note that $\mathbb{E}[A] = 0$. Thus,

$$= \mathbb{E}\Big[\Big(\sum_{k=1}^{K} \hat{w}_{tk}(\hat{\mu}_{tk} - \mathbb{E}[\hat{\mu}_{tk}])\Big)^2\Big] + \mathbb{E}\Big[\Big((\hat{w}_{tk} - w_{tk})\mathbb{E}[\hat{\mu}_{tk}] - w_{tk}\mu_{tk}P(|S_{tk}| = 0)\Big)^2\Big] \tag{147}$$

$$\leq \sum_{k=1}^{K} \hat{w}_{tk}^2 \mathrm{Var}[\hat{\mu}_{tk}] + \Big[\max_k C_k^{\mu_{tk}^2}\Big]\mathbb{E}\Big[\Big(w_{tk}P(|S_{tk}| = 0) + (\hat{w}_{tk} - w_{tk})\Big)^2\Big] \tag{148}$$

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

$$\mathrm{Var}[\hat{\mu}_{tk}] = \mathbb{E}[\hat{\mu}_{tk}^2] - \mathbb{E}[\hat{\mu}_{tk}]^2 \tag{149}$$

$$= P(|S_{tk}| > 0)(\mathbb{E}\Big[\frac{\sigma_{tk}^2}{|S_{tk}|}\Big||S_{tk}| > 0\Big] + \mu_{tk}^2) - P(|S_{tk}| > 0)^2\mu_{tk}^2 \tag{150}$$

$$= P(|S_{tk}| > 0)\mathbb{E}\Big[\frac{\sigma_{tk}^2}{|S_{tk}|}\Big||S_{tk}| > 0\Big] + P(|S_{tk}| > 0)(1 - P(|S_{tk}| > 0))\mu_{tk}^2 \tag{151}$$

$$= P(|S_{tk}| > 0)\mathbb{E}\Big[\frac{\sigma_{tk}^2}{|S_{tk}|}\Big||S_{tk}| > 0\Big] + P(|S_{tk}| > 0)P(|S_{tk}| = 0)\mu_{tk}^2 \tag{152}$$

$$\leq \mathbb{E}\Big[\frac{\sigma_{tk}^2}{|S_{tk}|}\Big||S_{tk}| > 0\Big] + P(|S_{tk}| = 0)\mu_{tk}^2 \tag{153}$$

$$\sum_{k=1}^{K} \hat{w}_{tk}^2 \mathrm{Var}[\hat{\mu}_{tk}] \leq \sum_{k=1}^{K} \hat{w}_{tk}^2 \Big(\mathbb{E}\Big[\frac{\sigma_{tk}^2}{|S_{tk}|}\Big||S_{tk}| > 0\Big] + P(|S_{tk}| = 0)\mu_{tk}^2\Big) \tag{154}$$

$$\leq \sum_{k=1}^{K} \Bigg(O\Big(p_{tk}^2\Big) + O\Big(\frac{p_{tk}^{3/2}}{\sqrt{N_1}}\Big)\Bigg)\Big(\mathbb{E}\Big[\frac{\sigma_{tk}^2}{|S_{tk}|}\Big||S_{tk}| > 0\Big] + P(|S_{tk}| = 0)\mu_{tk}^2\Big) \tag{155}$$

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

$$\leq \sum_{p_{<tk}>p^*_<}^{K} \left(O\big(p_{tk}^2\big)+O\left(\frac{p_{tk}^{3/2}}{\sqrt{N_1}}\right)\right)\left(\mathbb{E}\left[\frac{\sigma_{tk}^2}{|S_{tk}|}\bigg||S_{tk}|>0\right]+P(|S_{tk}|=0)\mu_{tk}^2\right)+\sum_{p_{<tk}\leq p^*_<}^{K} O\left(\frac{1}{\sqrt{N_{<t}}}\right) \tag{156}$$

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

$$\leq \sum_{p_{<tk}>p^*_<}^{K} \left(O\big(p_{tk}^2\big)+O\left(\frac{p_{tk}^{3/2}}{\sqrt{N_1}}\right)\right)\mathbb{E}\left[\frac{\sigma_{tk}^2}{|S_{tk}|}\bigg||S_{tk}|>0\right]+O\left(\frac{1}{N_1^2}\right)+O\left(\frac{N_1}{N_2^2}\right) \tag{157}$$

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

$$\leq \sum_{p_{<tk} > p_{<}^*}^{K} \left(O\big(p_{tk}^2\big) + O\bigg(\frac{p_{tk}^{3/2}}{\sqrt{N_1}}\bigg) \right) \left(\mathbb{E}\Big[\frac{\sigma_{tk}^2}{|S_{tk}|}\Big| |S_{tk}| > 0\Big] \right) + O\bigg(\frac{1}{N_1^2}\bigg) + O\bigg(\frac{N_1}{N_2^2}\bigg) \tag{158}$$

$$\leq \sum_{p_{<tk} > p_{<}^*}^{K} \left(\dots \right) \left(P(|S_{tk}| > F_{tk}) \mathbb{E}\Big[\frac{\sigma_{tk}^2}{|S_{tk}|}\Big| |S_{tk}| > 0\Big] + O\bigg(\frac{1}{N_1}\bigg) C_{tk}^{\sigma_{tk}^2} \right) + O\bigg(\frac{1}{N_1^2}\bigg) + O\bigg(\frac{N_1}{N_2^2}\bigg) \tag{159}$$

$$\leq \sum_{p_{<tk} > p_{<}^*}^{K} \left(O\big(p_{tk}^2\big) + O\bigg(\frac{p_{tk}^{3/2}}{\sqrt{N_1}}\bigg) \right) \left(\mathbb{E}\Big[\frac{\sigma_{tk}^2}{|S_{tk}|}\Big| |S_{tk}| > 0\Big] + O\bigg(\frac{1}{N_1}\bigg) C_{tk}^{\sigma_{tk}^2} \right) + O\bigg(\frac{1}{N_1^2}\bigg) + O\bigg(\frac{N_1}{N_2^2}\bigg) \tag{160}$$

$$\leq \sum_{p_{<tk} > p_{<}^*}^{K} \left(O\big(p_{tk}^2\big) + O\bigg(\frac{p_{tk}^{3/2}}{\sqrt{N_1}}\bigg) \right) \mathbb{E}\Big[\frac{\sigma_{tk}^2}{|S_{tk}|}\Big| |S_{tk}| > 0\Big] + O\bigg(\frac{1}{N_1^2}\bigg) + O\bigg(\frac{N_1}{N_2^2}\bigg) \tag{161}$$

$$\leq \sum_{p_{<tk} > p_{<}^*}^{K} \left(O\big(p_{tk}^2\big) + O\bigg(\frac{p_{tk}^{3/2}}{\sqrt{N_1}}\bigg) \right) \frac{\sigma_{tk}^2}{p_{tk} a_{tk} N_2 \left(1 - O\left(\sqrt{\frac{1}{p_{tk}(t-1)N_1}}\right) - O\left(\sqrt{\frac{\ln(1/\gamma)}{p_{tk}^{3/2} N_2}}\right) \right)} + O\bigg(\frac{1}{N_1^2}\bigg) + O\bigg(\frac{N_1}{N_2^2}\bigg) \tag{162}$$

$$\leq \sum_{p_{<tk} > p_{<}^*}^{K} \left(O\bigg(\frac{\sqrt{p_{tk}}}{N_2}\bigg) + O\bigg(\frac{1}{N_2\sqrt{N_1}}\bigg) \right) \left(1 + O\left(\sqrt{\frac{1}{p_{tk}(t-1)N_1}}\right) + O\left(\sqrt{\frac{\ln(1/\gamma)}{p_{tk}^{3/2} N_2}}\right) \right) + O\bigg(\frac{1}{N_1^2}\bigg) + O\bigg(\frac{N_1}{N_2^2}\bigg) \tag{163}$$

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 $\ln(1/\gamma) = \sqrt{N_1}$:

$$\leq \sum_{p_{<tk>} p^*_<}^{K} \left(O\left(\frac{1}{N_2\sqrt{N_1}}\right) + O\left(\frac{1}{N_2\sqrt{N_1}}\right)\right)\left(1 + O\left(\sqrt{\frac{N_1}{(t-1)N_1}}\right) + O\left(\sqrt{\frac{N_1^2}{N_2}}\right)\right) + O\left(\frac{1}{N_1^2}\right) + O\left(\frac{N_1}{N_2^2}\right) \tag{164}$$

$$\leq \sum_{p_{<tk>} p^*_<}^{K} O\left(\frac{1}{N_2\sqrt{N_1}}\right)\left(1 + O\left(\sqrt{\frac{N_1}{(t-1)N_1}}\right) + O\left(\sqrt{\frac{N_1^2}{N_2}}\right)\right) + O\left(\frac{1}{N_1^2}\right) + O\left(\frac{N_1}{N_2^2}\right) \tag{165}$$

$$\leq \sum_{p_{<tk>} p^*_<}^{K} O\left(\frac{1}{N_2\sqrt{N_1}}\right) + O\left(\frac{1}{N_2\sqrt{(t-1)N_1}}\right) + O\left(\frac{\sqrt{N_1}}{N_2\sqrt{N_2}}\right) + O\left(\frac{1}{N_1^2}\right) + O\left(\frac{N_1}{N_2^2}\right) \tag{166}$$

$$\leq O\left(\frac{1}{N_2\sqrt{N_1}}\right) + O\left(\frac{\sqrt{N_1}}{N_2\sqrt{N_2}}\right) + O\left(\frac{1}{N_1^2}\right) + O\left(\frac{N_1}{N_2^2}\right) \tag{167}$$

We will now upper bound the second term in the expression on (148). We apply Lemma 6.11 on (168):

$$\left[\max_k C_k^{\mu_{tk}^2}\right]\mathbb{E}\left[\left(w_{tk}P(|S_{tk}|=0) + (\hat{w}_{tk} - w_{tk})\right)^2\right] \leq \left[\max_k C_k^{\mu_{tk}^2}\right]\mathbb{E}\left[\left(O\left(\frac{1}{N_1}\right) + O\left(\frac{N_1}{N_2^2}\right) + O\left(\frac{1}{\sqrt{N_1}}\right)\right)^2\right] \tag{168}$$

$$\leq \left[\max_k C_k^{\mu_{tk}^2}\right]\mathbb{E}\left[\left(O\left(\frac{N_1}{N_2^2}\right) + O\left(\frac{1}{\sqrt{N_1}}\right)\right)^2\right] \tag{169}$$

$$\leq O\left(\frac{N_1^2}{N_2^4}\right) + O\left(\frac{\sqrt{N_1}}{N_2^2}\right) + O\left(\frac{1}{N_1}\right) \tag{170}$$

Finally, putting (167) and (170) together we have:

$$
\begin{aligned}
\mathbb{E}[(\hat{\mu}_t - \mu_t)^2] &= \mathbb{E}\Bigg[\Big(\sum_{k=1}^{K} \hat{w}_{tk}\hat{\mu}_{tk} - w_{tk}\mu_{tk}\Big)^2\Bigg] && (171)\\
&= \sum_{k=1}^{K} \hat{w}_{tk}^2 \mathrm{Var}[\hat{\mu}_{tk}] + \Big[\max_k C_k^{\mu_{tk}^2}\Big]\mathbb{E}\Big[\Big(w_{tk}P(|S_{tk}| = 0) + (\hat{w}_{tk} - w_{tk})\Big)^2\Big] && (172)\\
&\leq O\left(\frac{1}{N_2\sqrt{N_1}}\right) + O\left(\frac{\sqrt{N_1}}{N_2\sqrt{N_2}}\right) + O\left(\frac{1}{N_1^2}\right) + O\left(\frac{N_1}{N_2^2}\right) + O\left(\frac{N_1^2}{N_2^4}\right) + O\left(\frac{\sqrt{N_1}}{N_2^2}\right) + O\left(\frac{1}{N_1}\right) && (173)\\
&\leq O\left(\frac{1}{N_2\sqrt{N_1}}\right) + O\left(\frac{\sqrt{N_1}}{N_2\sqrt{N_2}}\right) + O\left(\frac{N_1}{N_2^2}\right) + O\left(\frac{N_1^2}{N_2^4}\right) + O\left(\frac{1}{N_1}\right) && (174)\\
&\leq O\left(\frac{1}{N_1}\right) + O\left(\frac{N_1}{N_2^2}\right) + O\left(\frac{1}{N_2\sqrt{N_1}}\right) && (175)
\end{aligned}
$$

7 Appendix

TODO

8 References

TODO