A Clipped Estimator

As mentioned in Section 5, the motivating equation guiding the design of our estimators is Eq. (2). This equation tells us that even when the statistics of the samples observed are governed by the distribution of (X, V, Y) under expert j, we can infer the mean of expert k. Such observations were made in [23, 27] in the context of best arm identification problems. Suppose we observe t samples under expert π_j. Guided by Eq. (2), one might come up with the following naive importance sampled estimator for the mean under expert k (μ_k):

$$\hat{\mu}_k = \frac{1}{t}\sum_{s=1}^{t} Y_j(s)\frac{\pi_k(V_j(s)|X_j(s))}{\pi_j(V_j(s)|X_j(s))}.$$

However, it is not possible to derive good confidence interval for the above estimator because even though the reward variable Y is bounded, the reweighing term $\pi_k(V_j(s)|X_j(s))/\pi_j(V_j(s)|X_j(s))$ can be unbounded and in some case heavy-tailed. The key idea is to come up with robust estimators that have good variance properties. One approach of controlling the variance of such estimators is to clip that the samples that are too large. This leads to the following clipped estimator [27]:

$$\hat{\mu}_k = \frac{1}{t}\sum_{s=1}^{t} Y_j(s)\frac{\pi_k(V_j(s)|X_j(s))}{\pi_j(V_j(s)|X_j(s))}\times \mathbb{1}\left(\frac{\pi_k(V_j(s)|X_j(s))}{\pi_j(V_j(s)|X_j(s))} \leq \eta_{kj}\right). \tag{9}$$

The clipping makes the estimator biased, however it helps in controlling the variance. The clipper level η_{kj} which depends on the relationship between π_k and π_j needs to be set carefully to control the bias-variance trade-off. In [27], it has been shown that if the log-divergence measure M_{kj} (defined in (2)) is bounded, then a good choice is $\eta_{kj} = 2\log(2/\epsilon)M_{kj}$, if we want an additive bias of at most $\epsilon(t)/2$ (Theorem 3 in [27]).

This idea can be generalized to estimating the mean of expert k, while observing samples from all the other experts. This leads to the clipped estimator in Eq. (3). Lemma 1 provides confidence guarantees for this estimator. The proof of this lemma follows from Theorem 4 in [27], but we include it here for completeness. In what follows, we will abbreviate $\mathbb{E}_{p_j(.)}[.]$ as $\mathbb{E}_j[.]$. In this section let $\hat{\mu}_k(t) = \hat{\mu}_k^c(t)$.

Proof of Lemma 1. Note that from Lemma 3 in [27] it follows that:

$$\mathbb{E}_j\left[\hat{\mu}_k(t)\right] \leq \mu_k \leq \mathbb{E}_j\left[\hat{\mu}_k(t)\right] + \frac{\epsilon(t)}{2}. \tag{10}$$

For the sake of analysis, let us consider the rescaled version $\bar{\mu}_k(t) = (Z_k(t)/t)\hat{\mu}_k(t)$ which can be written as:

$$\bar{\mu}_k(t) = \frac{1}{t}\sum_{j=0}^{N}\sum_{s\in\mathcal{T}_j}\frac{1}{M_{kj}}Y_j(s)\frac{\pi_k(V_j(s)|X_j(s))}{\pi_j(V_j(s)|X_j(s))} \times \mathbb{1}\left\{\frac{\pi_k(V_j(s)|X_j(s))}{\pi_j(V_j(s)|X_j(s))} \leq 2\log(2/\epsilon(t))M_{kj}\right\}. \tag{11}$$

Since $Y_j(s) \leq 1$, we have every random variable in the sum in (11) bounded by $2\log(2/\epsilon(t))$

Let, $\bar{\mu}_k = \mathbb{E}[\bar{\mu}_k(t)]$. Therefore by Chernoff bound, we have the following chain:

$$\begin{aligned}
&\mathbb{P}\left(|\bar{\mu}_k(t) - \bar{\mu}_k| \leq \delta\right) \leq 2\exp\left(-\frac{\delta^2 t}{8(\log(2/\epsilon(t)))^2}\right) \\
&\implies \mathbb{P}\left(|\bar{\mu}_k(t)\frac{t}{Z_k(t)} - \bar{\mu}_k\frac{t}{Z_k(t)}| \leq \delta\frac{t}{Z_k(t)}\right) \qquad (12)\\
&\leq 2\exp\left(-\frac{\delta^2 t}{8(\log(2/\epsilon(t)))^2}\right) \\
&\implies \mathbb{P}\left(|\hat{\mu}_k(t) - \hat{\mu}_k| \leq \delta\frac{t}{Z_k(t)}\right) \qquad (13)\\
&\leq 2\exp\left(-\frac{\delta^2 t}{8(\log(2/\epsilon(t)))^2}\right) \\
&\implies \mathbb{P}\left(|\hat{\mu}_k(t) - \hat{\mu}_k| \leq \delta\right) \qquad (14)\\
&\leq 2\exp\left(-\frac{\delta^2 t}{8(\log(2/\epsilon(t)))^2}\left(\frac{Z_k(t)}{t}\right)^2\right) \qquad (15)
\end{aligned}$$

Now we can combine Equations (12) and (10) to obtain:

$$\begin{aligned}
&\mathbb{P}\left(\mu_k - \delta - \epsilon(t)/2 \leq \hat{\mu}_k(t) \leq \mu_k + \delta\right) \\
&\geq 1 - 2\exp\left(-\frac{\delta^2 t}{8(\log(2/\epsilon(t)))^2}\left(\frac{Z_k(t)}{t}\right)^2\right)
\end{aligned}$$

□

Now, we will prove Theorem 1. Note that we will reindex the experts such that $0 = \Delta_{(1)} \leq \Delta_{(2)} \leq ... \leq \Delta_{(N)}$. Note that throughout this proof $U_k(t), \hat{\mu}_k(t)$ and $s_k(t)$ in Algorithm 1 are defined as in Equations (3) and (4) respectively. Before we proceed let us prove some key lemmas.

First we prove that with high enough probability the upper confidence bound estimate for the optimal expert k^* is greater than the true mean μ^*.

Lemma 3. *We have the following confidence bound at time t,*

$$\mathbb{P}\left(U_{k^*}(t) \leq \mu^*\right) \geq 1 - \frac{1}{t^2}.$$

Proof. We have the following chain,

$$
\begin{aligned}
&\mathbb{P}\left(U_{k^*}(t) \geq \mu^*\right) = \mathbb{P}\left(\hat{\mu}_{k^*}(t) \geq \mu^* - s_{k^*}(t)\right) \\
&= \mathbb{P}\left(\hat{\mu}_{k^*}(t) \geq \mu^* - \frac{3\beta(t)}{2}\right) \\
&\geq 1 - \frac{1}{t^2}.
\end{aligned}
$$

The last inequality is obtained by setting $\delta = \beta(t)$ and $\epsilon(t) = \beta(t)$ in Lemma 1. □

Next we prove that for a large enough time t, the UCB estimate of the k^{th} expert is less than that of μ^*.

Lemma 4. *We have the following confidence bound at time* $t > \frac{144M^2 \log^2(6/\Delta_k) \log T}{\Delta_k^2}$,

$$
\mathbb{P}\left(U_k(t) < \mu^*\right) \geq 1 - \frac{1}{t^2}.
$$

Proof. We have the following chain,

$$
\begin{aligned}
\mathbb{P}\left(U_k(t) > \mu^*\right) &= \mathbb{P}\left(\hat{\mu}_k(t) > \mu^* - \frac{3\beta(t)}{2}\right) \\
&\overset{(i)}{\leq} \mathbb{P}\left(\hat{\mu}_k(t) > \mu^* - \frac{\Delta_k}{2}\right) \\
&\overset{(ii)}{\leq} \mathbb{P}\left(\hat{\mu}_k(t) > \mu_k + \frac{\Delta_k}{2}\right) \\
&\overset{(iii)}{\leq} \frac{1}{t^2}.
\end{aligned}
$$

Here, (i) follows from the fact that $\frac{Z_k(t)}{t} \geq 1/M$, $t > \frac{144M^2 \log^2(6/\Delta_k) \log T}{\Delta_k^2}$ and the definition of $\beta(t)$. (ii) is by definition of Δ_k. Finally the concentration bound in (iii) follows from Lemma 1. □

Note that Lemmas 3 and 4 together imply,

$$
\mathbb{P}\left(k(t) = k\right) \leq \frac{2}{t^2} \tag{16}
$$

for $k \neq k^*$ and $t > \frac{144M^2 \log^2(6/\Delta_k) \log T}{\Delta_k^2}$.

Proof of Theorem 1. Let $T_k = \frac{144M^2 \log^2(6/\Delta_{(k)}) \log T}{\Delta_{(k)}^2}$ for $k = 2, .., N$. The regret of the algorithm can be bounded as,

$$
\begin{aligned}
&R(T) \leq \Delta_{(N)} T_N \\
&+ \sum_{k=0}^{N-3} \sum_{t=T_{N-k}}^{T_{N-k-1}} \Big(\Delta_{(N-k-1)} \mathbb{P}\left(\mathbb{1}\{k(t) \in \{(1), ..., (N-k-1)\}\}\right) \\
&+ \sum_{i=N-k}^{N} \Delta_{(i)} \mathbb{P}\left(\mathbb{1}\{k(t) = (i)\}\right) \Big) \\
&\leq \Delta_{(N)} T_N + \sum_{k=0}^{N-3} \Big(\Delta_{(N-k-1)} \left(T_{N-k-1} - T_{N-k}\right) \\
&+ \sum_{t=T_{N-k}}^{T_{N-k-1}} \sum_{i=N-k}^{N} \frac{2\Delta_{(i)}}{t^2} \Big) \qquad (17) \\
&\leq \sum_{k=0}^{N-3} \frac{144M^2 \log^2(6/\Delta_{(N-k-1)}) \log T}{\Delta_{(N-k-1)}} \left(1 - \frac{\gamma(\Delta_{(N-k-1)})}{\gamma(\Delta_{(N-k)})}\right) \\
&+ \Delta_{(N)} T_N + \frac{\pi^2}{3} \left(\sum_{i=2}^{N} \Delta_{(i)}\right) \\
&= \frac{144M^2 \log^2(6/\Delta_{(N)}) \log T}{\Delta_{(N)}} + \frac{\pi^2}{3} \left(\sum_{i=2}^{N} \Delta_{(i)}\right) \\
&+ \sum_{k=2}^{N-1} \frac{144M^2 \log^2(6/\Delta_{(k)}) \log T}{\Delta_{(k)}} \left(1 - \frac{\gamma(\Delta_{(k)})}{\gamma(\Delta_{(k+1)})}\right).
\end{aligned}
$$

Here, $\gamma(x) = \frac{x^2}{\log^2(6/x)}$.

□

B Median of Means Estimator

The median of means estimator is popular for estimating statistics under heavy-tailed distribution [13, 26]. We shall see that the median of means based estimator in Eq. (6) has good variance properties, when the chi-square divergence (Assumption 2) are bounded. Before proving Lemma 2, we will be establishing some intermediate results.

Lemma 5. *Consider the quantity* $\hat{\mu}_k^r(t)$ *in Eq. (5). The variance of this quantity is upper bounded as follows:*

$$
\mathrm{Var}[\hat{\mu}_k^r(t)] \leq \frac{1}{m} . \frac{1}{W_k(t)^2} \leq \frac{\sigma^2}{m}
$$

where $m = \lfloor t/l(t) \rfloor$.

Proof. We have the following chain,

$$
\begin{aligned}
&\mathrm{Var}[\hat{\mu}_k^r(t)] = \frac{1}{W_k(r,t)^2}\sum_{j=1}^{N}\sum_{s\in\mathcal{T}_j^{(r)}}\frac{1}{\sigma_{kj}^2}\mathrm{Var}\left(Y_j(s)\times\right.\\
&\left.\frac{\pi_k(V_j(s)|X(s))}{\pi_j(V_j(s)|X_j(s))}\right)\\
&\le \frac{1}{W_k(r,t)^2}\sum_{j=1}^{N}\sum_{s\in\mathcal{T}_j^{(r)}}\frac{1}{\sigma_{kj}^2}\mathrm{Var}\left(\frac{\pi_k(V_j(s)|X(s))}{\pi_j(V_j(s)|X_j(s))}\right)\\
&= \frac{m}{W_k(r,t)^2}\\
&= \frac{1}{m}\frac{1}{\left(\sum_{j=1}^{N}\frac{n_j(r,t)}{n(r,t).\sigma_{kj}}\right)^2}\\
&\le \frac{1}{m}.\frac{1}{W_k(t)^2} \qquad (18)
\end{aligned}
$$

□

Now, we can apply Chebyshev to conclude that for all $r \in [l(t)]$,

$$
\mathbb{P}\left(|\hat{\mu}_k^r(t)-\mu_k| \le \frac{1}{W_k(t)}\sqrt{\frac{4}{m}}\right) \ge \frac{3}{4}. \qquad (19)
$$

Now we will prove Lemma 2.

Proof of Lemma 2. In light of Eq. (19), the probability that the median is not within distance $\frac{1}{W_k(t)}\sqrt{\frac{4}{m}}$ of μ_k is bounded as,

$$
\begin{aligned}
&\mathbb{P}\left(|\hat{\mu}_k^m(t)-\mu_k| > \frac{1}{W_k(t)}\sqrt{\frac{4}{m}}\right)\\
&\le \mathbb{P}\left(\mathrm{Bin}(l(t),1/4) > l(t)/2\right) \le \delta(t).
\end{aligned}
$$

This concludes the proof. □

Note that we will re-index the experts such that $0 = \Delta_{(1)} \le \Delta_{(2)} \le ... \le \Delta_{(N)}$. Note that throughout this proof $U_k(t), \hat{\mu}_k(t)$ and $s_k(t)$ in Algorithm 1 are defined as in Equations (6) and (7) respectively. Before we proceed let us prove some key lemmas. Now we prove lemmas analogous to Lemmas 4 and 3.

Lemma 6. *We have the following confidence bound at time t,*

$$
\mathbb{P}\left(U_{k^*}(t) \le \mu^*\right) \ge 1 - \frac{1}{t^2}.
$$

The proof follows directly from Lemma 2.

Lemma 7. *We have the following confidence bound at time $t > \frac{256\sigma^2\log T}{\Delta_k^2}$,*

$$
\mathbb{P}\left(U_k(t) < \mu^*\right) \ge 1 - \frac{1}{t^2}.
$$

Proof. We have the following chain,

$$
\begin{aligned}
\mathbb{P}\left(U_k(t) > \mu^*\right) &= \mathbb{P}\left(\hat{\mu}_k(t) > \mu^* - \frac{1}{W_k(t)}\sqrt{\frac{64\log t}{t}}\right)\\
&\overset{(i)}{\le} \mathbb{P}\left(\hat{\mu}_k(t) > \mu^* - \frac{\Delta_k}{2}\right)\\
&\overset{(ii)}{\le} \mathbb{P}\left(\hat{\mu}_k(t) > \mu_k + \frac{\Delta_k}{2}\right)\\
&\overset{(iii)}{\le} \frac{1}{t^2}.
\end{aligned}
$$

Here, (i) follows from the fact that $\frac{1}{W_k(t)} \le \sigma$ and $t > \frac{256\sigma^2\log T}{\Delta_k^2}$. (ii) is by definition of Δ_k. Finally the concentration bound in (iii) follows from Lemma 2. □

Note that Lemma 7 and 6 together imply that,

$$
\mathbb{P}\left(k(t) = k\right) \le \frac{2}{t^2} \qquad (20)
$$

for $k \ne k^*$ and $t > \frac{256\sigma^2\log T}{\Delta_k^2}$.

Proof of Theorem 2. Let $T_k = \frac{256\sigma^2\log T}{\Delta_{(k)}^2}$ for $k = 2,..,N$. The regret of the algorithm can be bounded as,

$$
\begin{aligned}
&R(T) \leq \Delta_{(N)} T_N \\
&+ \sum_{k=0}^{N-3} \sum_{t=T_{N-k}}^{T_{N-k-1}} \left(\mathbb{P}\left(\mathbb{1}\{k(t) \in \{(1), \ldots, (N-k-1)\}\}\right)\right. \\
&\left.\times \Delta_{(N-k-1)} + \sum_{i=N-k}^{N} \Delta_{(i)} \mathbb{P}\left(\mathbb{1}\{k(t) = (i)\}\right)\right) \\
&\leq \Delta_{(N)} T_N + \sum_{k=0}^{N-3} \left(\Delta_{(N-k-1)} \left(T_{N-k-1} - T_{N-k}\right)\right. \\
&\left.+ \sum_{t=T_{N-k}}^{T_{N-k-1}} \sum_{i=N-k}^{N} \frac{2\Delta_{(i)}}{t^2}\right) \qquad (21) \\
&\leq \sum_{k=0}^{N-3} \frac{256\sigma^2 \log T}{\Delta_{(N-k-1)}} \left(1 - \frac{\Delta^2_{(N-k-1)}}{\Delta^2_{(N-k)}}\right) \\
&+ \Delta_{(N)} T_N + \frac{\pi^2}{3} \left(\sum_{i=2}^{N} \Delta_{(i)}\right) \\
&= \frac{256\sigma^2 \log T}{\Delta_{(N)}} + \frac{\pi^2}{3} \left(\sum_{i=2}^{N} \Delta_{(i)}\right) \\
&+ \sum_{k=2}^{N-1} \frac{256\sigma^2 \log T}{\Delta_{(k)}} \left(1 - \frac{\Delta^2_{(k)}}{\Delta^2_{(k+1)}}\right).
\end{aligned}
$$

□

C Instance Dependent Terms

In this section we devote our attention to the instance dependent terms in Theorems 1 and 2. We will first prove Corollary 1.

Proof of Corollary 1. We will prove the two statements about the two estimators separately,

(i) Going back to Lemma 4 in the proof of Theorem 1, we get that,

$$
\mathbb{P}(U_k(t) < \mu^*) \geq 1 - \frac{1}{t^2}
$$

when $t > \frac{144M^2 \log^2(6/\Delta_{(2)}) \log T}{\Delta_k^2}$. This simply follows from the fact that $\Delta_{(2)}$ is the smallest gap. Therefore, the chain leading to Eq. (17) follows with the new definition of $T_k = \frac{144M^2 \log^2(6/\Delta_{(2)}) \log T}{\Delta^2_{(k)}}$. Hence, the regret of Algorithm 1 under estimator (3) is bounded as follows:

$$
\begin{aligned}
&R(T) \leq \frac{144M^2 \log^2(6/\Delta_{(2)}) \log T}{\Delta_{(2)}} \left(1\right. \\
&\left.+ \sum_{k=2}^{N-1} \left(1 - \frac{\Delta^2_{(k)}}{\Delta^2_{(k+1)}}\right)\right) \\
&+ \frac{\pi^2}{3} \left(\sum_{i=2}^{N} \Delta_{(i)}\right) \\
&= \frac{144\lambda(\boldsymbol{\mu}) M^2 \log^2(6/\Delta_{(2)}) \log T}{\Delta_{(2)}} + \frac{\pi^2}{3} \left(\sum_{i=2}^{N} \Delta_{(i)}\right).
\end{aligned}
\qquad (22)
$$

We can analyze the same terms in an alternate manner. From Eq. (17) in the proof of Theorem 1, it follows that the regret of Algorithm 1 under the clipped estimator is bounded by,

$$
R(T) \leq \Delta_{(N)} T_2 + \frac{\pi^2}{3} \left(\sum_{i} \Delta_{(i)}\right).
$$

Using the definition of T_2 in (17) we obtain:

$$
R(T) \leq \frac{144M^2 \log^2(6/\Delta_{(2)}) \log T}{\Delta_{(2)}} \times \frac{1}{\Delta_{(2)}}.
$$

Combining the above equation with (22) we get the desired result.

(ii) Theorem 2 immediately implies that

$$
R(T) \leq \frac{256\lambda(\boldsymbol{\mu})\sigma^2 \log T}{\Delta_{(2)}} + \frac{\pi^2}{3} \left(\sum_{i=2}^{N} \Delta_{(i)}\right),
$$

for the median of means estimator.

We can alternately analyze the regret as follows. From Eq. (21) in the proof of Theorem 2, it follows that the regret of Algorithm 1 under the median of means estimator is bounded by,

$$
R(T) \leq \Delta_{(N)} T_2 + \frac{\pi^2}{3} \left(\sum_{i} \Delta_{(i)}\right).
$$

Using the definition of T_2 in (21) we obtain:

$$
R(T) \leq \frac{256\sigma^2 \log T}{\Delta_{(2)}} \times \frac{1}{\Delta_{(2)}}.
$$

Combining the equations above we get the desired result.

□

Now we will work under the assumption that the gaps in the means of the experts are generated according to the generative model in Corollary 2.

Proof of Corollary 2. In light of Corollary 1, we just need to prove that $\mathbb{E}_{p_\Delta}[\lambda(\boldsymbol{\mu})] = O(\log N)$.

Now, we will assume that $\{\Delta_{(i)}\}$ for $i = 3, ..., N$, are order statistics of $N - 2$ i.i.d uniform r.vs over the interval $[\Delta_{(2)}, 1]$.

Note that by Jensen's we have the following:

$$1 - \mathbb{E}\left[\frac{\Delta^2_{(k)}}{\Delta^2_{(k+1)}}\right] \leq 1 - \mathbb{E}\left[\frac{\Delta_{(k)}}{\Delta_{(k+1)}}\right]^2. \tag{23}$$

Let $X = \Delta_{(k)}$ and $Y = \Delta_{(k+1)}$ for $k \geq 3$. The joint pdf of X, Y is given by,

$$f(x, y) = \frac{(N-2)!}{(k-1)!(N-3-k)!}\left(\frac{x - \Delta_{(2)}}{1 - \Delta_{(2)}}\right)^{k-1} \times \left(1 - \frac{y - \Delta_{(2)}}{1 - \Delta_{(2)}}\right)^{N-3-k} \frac{1}{(1 - \Delta_{(2)})^2}.$$

Therefore, we have the following chain:

$$\begin{aligned}
\mathbb{E}\left[\frac{X}{Y}\right] &= \int_{y=\Delta_{(2)}}^{1}\int_{x=\Delta_{(2)}}^{y} \frac{x}{y}\frac{(N-2)!}{(k-1)!(N-3-k)!} \times \\
&\left(\frac{x - \Delta_{(2)}}{1 - \Delta_{(2)}}\right)^{k-1}\left(1 - \frac{y - \Delta_{(2)}}{1 - \Delta_{(2)}}\right)^{N-3-k}\frac{1}{(1 - \Delta_{(2)})^2}dxdy \\
&= \int_{b=0}^{1}\int_{a=0}^{b}\frac{(1 - \Delta_{(2)})a + \Delta_{(2)}}{(1 - \Delta_{(2)})b + \Delta_{(2)}}\frac{(N-2)!}{(k-1)!(N-3-k)!} \times \\
&(a)^{k-1}(1 - b)^{N-3-k}\,dadb \\
&\geq \int_{b=0}^{1}\int_{a=0}^{b}\frac{a}{b}\frac{(N-2)!}{(k-1)!(N-3-k)!} \times \\
&(a)^{k-1}(1 - b)^{N-3-k}\,dadb \\
&= \frac{k}{k+1}.
\end{aligned}$$

Combining this with Eq (23) yields,

$$\begin{aligned}
&\mathbb{E}_{p_\Delta}\left[\sum_{k=2}^{N-1}\left(1 - \frac{\Delta^2_{(k)}}{\Delta^2_{(k+1)}}\right)\right] \\
&\leq 1 + \sum_{k=3}^{N-1}\left(1 - \frac{k^2}{(k+1)^2}\right) \\
&= 1 + \sum_{k=3}^{N-1}\left(\frac{2k+1}{(k+1)^2}\right) \\
&\leq 1 + \sum_{k=3}^{N-1}\frac{2}{k+1} \qquad (24) \\
&\leq 1 + 2\log N.
\end{aligned}$$

□

D More on Empirical Results

In this section we provide more details about our empirical results under the following sub-headings.

Training of Stochastic Experts: In Algorithm 2, new experts are added before starting a new batch. These stochastic experts are classifying functions trained using cost-sensitive classification oracles on data observed so far, which uses the ideas in [8]. The key idea is to reduce the cost-sensitive classification problem into importance weighted classification, which can be solved using binary classifiers by providing weights to each samples. Suppose a context x is observed and Algorithm 2 chooses an expert π_i and draws an arm a from the conditional distribution $\pi_i(V|x)$. Suppose the reward observed is $r(a)$. Then the training sample (x, a) with a sample weight of $r(a)/\pi_i(a|x)$ is added to the dataset for training the next batch of experts. It has been shown that this importance weighing yields *good* classification experts. These classifiers can provide confidence scores for arms, given a context and hence can serve as stochastic experts. 4 different experts are added at the beginning of each batch, out of which three are trained by XgBoost as base-classifier while one is trained by logistic regression. Diversity is maintained among the experts added by training them on bootstrapped versions of the data observed so far, and also through selecting different hyper-parameters. Note that the parameter selection scheme is not tuned per dataset, but is held fixed for all three datasets.

Estimating Divergence Parameters: Both our divergence metrics M_{ij}'s and σ_{ij}'s can be estimated from data observed so far, during a run of Algorithm 2. These divergences do not depend on the arm chosen, but only on the context distribution and the conditional distributions encoded by the expert. Therefore, they can be easily estimated from data observed. Suppose, n contexts have been observed so far $\{x_1, ..., x_n\}$. We are interested in estimating σ_{ij} that is the chi-square divergence between π_i and π_j. An estimator for this would be the empirical mean $(1/n) \times \sum_{k=1}^{n} D_{f_2}(\pi_i(.|x_k)\|\pi_j(.|x_k))$. Note that the distribution over the arms $\pi_j(.|x_k)$ is nothing but the confidence scores observed through evaluation of the classifying oracle π_j on the features/context x_i. In order to be robust, we use the median of means estimator instead of the simple empirical mean for estimating the divergences.

Empirical Analysis of Instance Dependent terms: In this section we empirically validate that our instance dependent terms in Theorem 1 and 2 are indeed much smaller compared to corresponding terms in the UCB-1 [5] regret bounds, even in real problem where our generative assumptions do not hold. In order to showcase this, we plot the instance-dependent term in Theorem 2 which is given by,

$$\sum_{k=2}^{N-1} \frac{1}{\Delta_{(k)}} \left(1 - \frac{\Delta_{(k)}^2}{\Delta_{(k+1)}^2}\right) + \frac{1}{\Delta_{(N)}} \tag{25}$$

along with the corresponding term in UCB-1 bounds given by,

$$\sum_{k=2}^{N} \frac{1}{\Delta_{(k)}}, \tag{26}$$

as the number of stochastic experts grow in the stream dataset experiments in Section 7. The true means of the experts have been estimated in hindsight over the whole dataset. The plot is shown in Fig. 3. It can be observed that the term in the bounds of D-UCB grows at a much slower pace, and in fact stops increasing with the number of experts after a certain point.

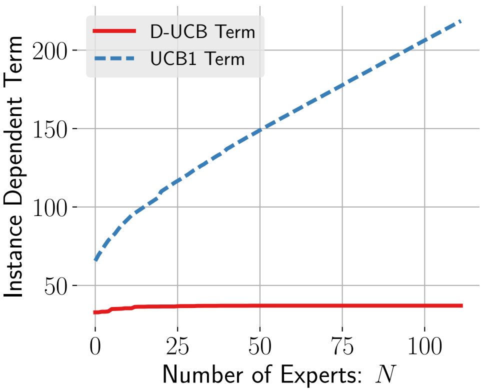

Figure 3: We plot the instance-dependent terms from D-UCB bounds (the term in Theorem 2 involving the gaps (25)) and that of UCB-1 bounds (26) as the number of experts grows in the stream analytics dataset. It can be observed that the instance dependent term from D-UCB grows at a much slower pace with the number of experts, and in fact stops increasing after a certain point.