Journal of Machine Learning Research 14 (2013) 601-623 Submitted 10/11; Revised 11/12; Published 2/13

Optimal Discovery with Probabilistic Expert Advice: Finite Time Analysis and Macroscopic Optimality

Sébastien Bubeck SBUBECK@PRINCETON.EDU
Department of Operations Research and Financial Engineering
Princeton University
Princeton, NJ, 08544, USA

Damien Ernst DERNST@ULG.AC.BE
Department of Electrical Engineering and Computer Science
University of Liège, Institut Montefiore, B28
B-4000 Liège, Belgium

Aurélien Garivier AURELIEN.GARIVIER@MATH.UNIV-TOULOUSE.FR
Institut de Mathématiques de Toulouse
Université Paul Sabatier
118, route de Narbonne
F-31062 Toulouse Cedex 9, France

Editor: Nicolo Cesa-Bianchi

Abstract

We consider an original problem that arises from the issue of security analysis of a power system and that we name optimal discovery with probabilistic expert advice. We address it with an algorithm based on the optimistic paradigm and on the Good-Turing missing mass estimator. We prove two different regret bounds on the performance of this algorithm under weak assumptions on the probabilistic experts. Under more restrictive hypotheses, we also prove a macroscopic optimality result, comparing the algorithm both with an oracle strategy and with uniform sampling. Finally, we provide numerical experiments illustrating these theoretical findings.

Keywords: optimal discovery, probabilistic experts, optimistic algorithm, Good-Turing estimator, UCB

1. Introduction

In this paper we consider the following problem: Let $\mathcal{X}$ be a set, and $A \subset \mathcal{X}$ be a set of interesting elements in $\mathcal{X}$. One can access $\mathcal{X}$ only through requests to a finite set of probabilistic experts. More precisely, when one makes a request to the i^{th} expert, the latter draws independently at random a point from a fixed probability distribution P_i over $\mathcal{X}$. One is interested in discovering rapidly as many elements of A as possible, by making sequential requests to the experts.

1.1 Motivation

The original motivation for this problem arises from the issue of real-time security analysis of a power system. This problem often amounts to identifying in a set of credible contingencies those that may indeed endanger the security of the power system and perhaps lead to a system collapse with catastrophic consequences (e.g., an entire region, country may be without electrical power for

hours). Once those dangerous contingencies have been identified, the system operators usually take preventive actions so as to ensure that they could mitigate their effect on the system in the likelihood they would occur. Note that usually, the dangerous contingencies are very rare with respect to the non dangerous ones. A straightforward approach for tackling this security analysis problem is to simulate the power system dynamics for every credible contingency so as to identify those that are indeed dangerous. Unfortunately, when the set of credible contingencies contains a large number of elements (say, there are more than 10^5 credible contingencies) such an approach may not possible anymore since the computational resources required to simulate every contingency may excess those that are usually available during the few (tens of) minutes available for the real-time security analysis. One is therefore left with the problem of identifying within this short time-frame a maximum number of dangerous contingencies rather than all of them. The approach proposed in Fonteneau-Belmudes (2012) and Fonteneau-Belmudes et al. (2010) addresses this problem by building first very rapidly what could be described as a probability distribution P over the set of credible contingencies that points with significant probability to contingencies which are dangerous. Afterwards, this probability distribution is used to draw the contingencies to be analyzed through simulations. When the computational resources are exhausted, the approach outputs the contingencies found to be dangerous. One of the main shortcoming of this approach is that usually P points only with a significant probability to a few of the dangerous contingencies and not all of them. This in turn makes this probability distribution not more likely to generate after a few draws new dangerous contingencies than for example a uniform one. The dangerous contingencies to which P points to with a significant probability depend however strongly on the set of (sometimes arbitrary) engineering choices that have been made for building it. One possible strategy to ensure that more dangerous contingencies can be identified within a limited budget of draws would therefore be to consider $K > 1$ sets of engineering choices to build K different probability distributions P_1, P_2, …, P_K and to draw the contingencies from these K distributions rather than only from a single one. This strategy raises however an important question to which this paper tries to answer: how should the distributions be selected for being able to generate with a given number of draws a maximum number of dangerous contingencies? We consider the specific case where the contingencies are sequentially drawn and where the distribution selected for generating a contingency at one instant can be based on the past distributions that have been selected, the contingencies that have been already drawn and the results of the security analyses (dangerous/non dangerous) for these contingencies. This corresponds exactly to the optimal discovery problem with expert advice described above. We believe that this framework has many other possible applications, such as for example web-based content access.

1.2 Setting and Notation

In this paper we restrict our attention to finite or countably infinite sets $\mathcal{X}$. We denote by K the number of experts. For each $i \in \{1,\dots,K\}$, we assume that $(X_{i,n})_{n\geq 1}$ are random variables with distribution P_i such that the $(X_{i,n})_{i,n}$ are independent. Sequential discovery with probabilistic expert advice can be described as follows: at each time step $t \in \mathbb{N}^*$, one picks an index $I_t \in \{1,\dots,K\}$, and one observes $X_{I_t,n_{I_t,t}}$, where

$$n_{i,t} = \sum_{s\leq t} \mathbb{1}\{I_s = i\} \, .$$

The goal is to choose the $(I_t)_{t\geq 1}$ so as to observe as many elements of A as possible in a fixed horizon t, that is to maximize the number of interesting items found after t requests

$$F(t) = \sum_{x\in A} \mathbb{1}\left\{x \in \{X_{1,1},\dots,X_{1,n_{1,t}},\dots,X_{K,1},\dots,X_{K,n_{K,t}}\}\right\}. \tag{1}$$

Note in particular that it is of no interest to observe twice the same same element of A. The index I_{t+1} may be chosen according to past observations: it is a (possibly randomized) function of $(I_1, X_{I_1,1},\dots,I_t,X_{I_t,n_{I_t,t}})$.

An easier quantity to analyze than the number of interesting items found $F(t)$ is the waiting time $T(\lambda)$, $\lambda \in (0,1)$, which is the time at which the strategy has a missing mass of interesting items smaller than λ on every experts, that is

$$T(\lambda) = \inf\left\{t : \forall i \in \{1,\dots,K\}, P_i(A \setminus \{X_{1,1},\dots,X_{1,n_{1,t}},\dots,X_{K,1},\dots,X_{K,n_{K,t}}\}) \leq \lambda\right\}. \tag{2}$$

While we shall derive a general strategy that can be used without any assumption on the probabilistic experts, for the mathematical analysis of the waiting time $T(\lambda)$ we make the following assumption:

(i) non-intersecting supports: $A \cap \mathrm{supp}(P_i) \cap \mathrm{supp}(P_j) = \emptyset$ for $i \neq j$.

Furthermore we will also derive some results under the following more restrictive assumptions:

(ii) finite supports with the same cardinality: $|\mathrm{supp}(P_i)| = N, \forall i \in \{1,\dots,K\}$,

(iii) uniform distributions: $P_i(x) = \frac{1}{N}, \forall x \in \mathrm{supp}(P_i), \forall i \in \{1,\dots,K\}$.

1.3 Contribution and Content of the Paper

This paper contains the description of a generic algorithm for the optimal discovery problem with probabilistic expert advice, and a theoretical analysis of its properties. In Section 2, we first depict our strategy, termed Good-UCB. This algorithm relies on the *optimistic paradigm*, which led to the UCB (Upper Confidence Bound) algorithm for multi-armed bandits, see Auer et al. (2002) and Garivier and Cappé (2011). It relies also on a finite-time analysis of the Good-Turing estimator for the missing mass. We also derive in Section 2 two different regret bounds under the non-intersecting assumption (i): we first show that $F^{UCB}(t)$ (the number of interesting items found by Good-UCB) is larger than $F^*(t)$ (the number of interesting items found by an oracle strategy), up to a term of order $\sqrt{Kt\log(t)}$. We argue that such a bound does not capture all the fine properties of Good-UCB: indeed, on the contrary to the multi-armed bandit problem, here the regret $F^*(t) - F(t)$ remains bounded for any reasonable strategy. This can be understood as a *restoring property* of the game: if a policy makes a sub-optimal choice at some given time t, then in the future it will have better opportunities than the optimal policy. This key feature of our problem prevents the regret from growing too much. To analyze this phenomenon, we complete our first bound by a second regret analysis—the main result of the paper—which states roughly that with high probability, $T_{UCB}(\lambda)$ (the waiting time for the strategy Good-UCB) is *uniformly* (in λ) smaller than $T^*(\lambda')$ (the smallest possible waiting time), for some λ' close to λ and up to a small additional term, see Theorem 5 for a more precise statement. We emphasize that these regret bounds are both completely distribution-free and explicit.

In Section 3 we propose to investigate the behavior of Good-UCB in a *macroscopic limit* sense, that is we make assumptions [(i), (ii), (iii)] and we consider the limit when the size of the set $\mathcal{X}$ grows to infinity while maintaining a constant proportion of interesting items. In this scenario we show that Good-UCB is macroscopically optimal, in the sense that the normalized waiting time of Good-UCB tends to the normalized smallest possible waiting time. We also derive a formula for this latter quantity and we show that it is equal to $\sum_{i:q_i>\lambda} \log \frac{q_i}{\lambda}$, where q_i is the limiting proportion of interesting items on expert i. This macroscopic limit also allows to easily assess the performance of different strategies, and we show that for example the normalized waiting time of uniform sampling tends to $K \max_{1\leq i\leq K} \log \frac{q_i}{\lambda}$, which proves that this strategy is macroscopically suboptimal, unless all experts have the same number of interesting items.

Finally, Section 4 reports experimental results that show that the Good-UCB algorithm performs very well, even in a setting where assumptions (i), (ii) and (iii) are not satisfied. The appendix contains some technical proofs, together with a more detailed discussion on oracle strategies in the macroscopic limit and on the relation between the waiting time T defined in (2) and the number of items found F defined in (1), proving in particular that optimality in terms of waiting time is equivalent to optimality in terms of number of items found.

2. The Good-UCB Algorithm

We describe here the Good-UCB strategy. This algorithm is a sequential method estimating at time t, for each expert $i \in \{1,\dots,K\}$, the total probability of the interesting items that remain to be discovered through requests to expert i. This estimation is done by adapting the so-called Good-Turing estimator for the missing mass. Then, instead of simply using the distribution with highest estimated missing mass, which proves hazardous, we make use of the *optimistic paradigm*—see Bubeck and Cesa-Bianchi (2012, Chapter 2, and references therein)—a heuristic principle well-known in reinforcement learning, which entails to prefer using an *upper-confidence bound* (UCB) of the missing mass instead. At a given time step, the Good-UCB algorithm simply makes a request to the expert with highest upper-confidence bound on the missing mass at this time step.

We start in Section 2.1 with the Good-Turing estimator and a brief study of its concentration properties. Then we describe precisely the Good-UCB strategy in Section 2.2. Next we proceed to the theoretical analysis of Good-UCB and we start in Section 2.3 where we describe an oracle strategy (that we shall use as a comparator) that we prove to be optimal under assumption (i). In Section 2.4 we show that one can obtain a standard regret bound of order $\sqrt{t}$ when one compares the number of items $F^{UCB}(t)$ found by Good-UCB to the number of items $F^*(t)$ found by the oracle. This bound is not completely satisfactory (as we explain in Section 2.4), and our main result—a 'non-linear' regret bound—is proved in Section 2.5.

2.1 Estimating the Missing Mass

Our algorithm relies on an estimation at each step of the probability of obtaining a new interesting item by making a request to a given expert. A similar issue was addressed by I. Good and A. Turing as part of their efforts to crack German ciphers for the Enigma machine during World War II. In this subsection, we describe a version of the Good-Turing estimator adapted to our problem. Let Ω be a discrete set, and let A be a subset of interesting elements of Ω. Assume that $X_1,\dots,X_n$ are elements

of Ω drawn independently under the same distribution P, and define for every $x \in \Omega$:

$$O_n(x) = \sum_{m=1}^{n} \mathbb{1}\{X_m = x\}, \quad Z_n(x) = \mathbb{1}\{O_n(x) = 0\}, \quad U_n(x) = \mathbb{1}\{O_n(x) = 1\} .$$

Let $R_n = \sum_{x \in A} Z_n(x) P(x)$ denote the missing mass of the interesting items, and let $U_n = \sum_{x \in A} U_n(x)$ be the number of elements of A that have been seen exactly once (in linguistics, they are often called *hapaxes*). The idea of the Good-Turing estimator—see Good (1953), see also McAllester and Schapire (2000); Orlitsky et al., and references therein—is to estimate the (random) "missing mass" R_n, which is the total probability of all the interesting items that do not occur in the sample $X_1, \ldots, X_n$, by the "fraction of hapaxes $\hat{R}_n = U_n/n$. This estimator is well-known in linguistics, for instance in order to estimate the number of words in some language, see Gale and Sampson (1995). We shall use the following tight bound on the estimation error. We emphasize the fact that the following bound holds true *independently of the underlying distribution P*.

Proposition 1 *With probability at least* $1 - \delta$*,*

$$\hat{R}_n - \frac{1}{n} - (1+\sqrt{2})\sqrt{\frac{\log(4/\delta)}{n}} \leq R_n \leq \hat{R}_n + (1+\sqrt{2})\sqrt{\frac{\log(4/\delta)}{n}} .$$

Proof For self-containment, we first show that $\mathbb{E}R_n - \mathbb{E}\hat{R}_n \in \left[-\frac{1}{n}, 0\right]$; this result is well known, see for example Theorem 1 in McAllester and Schapire (2000):

$$\begin{aligned}
\mathbb{E}R_n - \mathbb{E}\hat{R}_n &= \sum_{x \in A} \left[P(x)\left(1 - P(x)\right)^n - \frac{1}{n} \times nP(x)\left(1 - P(x)\right)^{n-1} \right] \\
&= -\frac{1}{n} \sum_{x \in A} P(x) \times nP(x)\left(1 - P(x)\right)^{n-1} \\
&= -\frac{1}{n} \mathbb{E}\left[\sum_{x \in A} P(x) U_n(x) \right] \in \left[-\frac{1}{n}, 0 \right] .
\end{aligned}$$

Next we apply the inequality of McDiarmid (1989) to $\hat{R}_n$ as follows. The random variable $\hat{R}_n$ is a function of the independent observations $X_1, \ldots, X_n$ such that, denoting $\hat{R}_n = f(X_1, \ldots, X_n)$, modifying just one observation has limited impact: $\forall l \in \{1, \ldots, n\}, \forall (x_1, \ldots, x_n, x_l') \in \Omega^{n+1}$,

$$\left| f(x_1, \ldots, x_n) - f(x_1, \ldots, x_{l-1}, x_l', x_{l+1}, \ldots, x_n) \right| \leq \frac{2}{n}.$$

Thus one gets that, with probability at least $1 - \delta$,

$$\left| \hat{R}_n - \mathbb{E}[\hat{R}_n] \right| \leq \sqrt{\frac{2\log(2/\delta)}{n}} .$$

Finally we extract the following result from Theorem 10 and Theorem 16 in McAllester and Ortiz (2003): with probability at least $1 - \delta$,

$$\left| R_n - \mathbb{E}[R_n] \right| \leq \sqrt{\frac{\log(2/\delta)}{n}} .$$

which concludes the proof. ■

2.2 The Good-UCB Algorithm

Following the example of the well-known Upper-Confidence Bound procedure for multi-armed bandit problems, we propose Algorithm 1, which we call *Good-UCB* in reference to the estimator it relies on. For each arm $i \in \{1,\dots,K\}$, the index at time t of Good-UCB corresponds to the estimate

$$\hat{R}_{i,n_{i,t-1}} = \frac{1}{n_{i,t-1}} \sum_{x \in A} \mathbb{1}\left\{ \sum_{s=1}^{n_{i,t-1}} \mathbb{1}\{X_{i,s} = x\} = 1 \text{ and } \sum_{j=1}^{K} \sum_{s=1}^{n_{j,t-1}} \mathbb{1}\{X_{j,s} = x\} = 1 \right\}$$

of the missing mass

$$\sum_{x \in A \setminus \left\{ X_{I_1,n_{I_1,1}},\dots,X_{I_{t-1},n_{I_{t-1},t-1}} \right\}} P_i(x) \tag{3}$$

inflated by a confidence bonus of order $\sqrt{\log(t)/n_{i,t-1}}$. Good-UCB relies on a tuning parameter C which is discussed below.

Algorithm 1 Good-UCB

For $1 \le t \le K$ choose $I_t = t$.
for $t \ge K+1$ **do**
  Choose $I_t = \arg\max_{1 \le i \le K} \left\{ \hat{R}_{i,n_{i,t-1}} + C\sqrt{\frac{\log(4t)}{n_{i,t-1}}} \right\}$
  Observe X_t distributed as P_{I_t} and update the missing mass estimates accordingly
end for

The Good-UCB algorithm is designed to work without any assumption on the probabilistic experts. However for the analysis we shall make the non-intersecting supports assumption (i). Indeed without this assumption the missing mass of a given expert i depends explicitly on the outcomes of *all* requests (and not only requests to expert i), see (3), which makes the analysis significantly more difficult. On the other hand under assumption (i) one can define the missing mass of expert i after n pulls without any reference to the other arms, and it takes the following simple form:

$$R_{i,n} = \sum_{x \in A \setminus \{X_{i,1},\dots,X_{i,n}\}} P_i(x)\,. \tag{4}$$

Note that while the theoretical analysis will be carried out under assumption (i), we show in Section 4 that Good-UCB performs well in practice even when this assumption is not met.

2.3 The Closed-loop Oracle Policy

In this section we define a policy that we shall use as a benchmark to study the properties of Good-UCB. We assume hereafter that assumption (i) is satisfied (in particular we shall use the notation defined in (4)). The Oracle Closed-Loop policy, denoted OCL in the following, makes a request at time t to the expert

$$I_t^* = \arg\max_{1 \le i \le K} R_{i,n_{i,t-1}^*}\,, \text{ where } n_{i,t}^* = \sum_{s=1}^{t} \mathbb{1}\{I_s^* = i\}\,.$$

In words, OCL greedily selects the expert that maximizes the probability of finding a new interesting item. The next lemma shows that this greedy procedure is in fact optimal (in expectation) under assumption (i). The proof is given in the appendix.

For any given policy π, let $F^\pi(t)$ be the number of items found at time t with π, I_t^π be the expert chosen by π at time t, and $n_{i,t}^\pi = \sum_{s=1}^t \mathbb{1}\{I_s^\pi = i\}$ be the number of requests made by π to expert i up to time t.

Lemma 2 *Let π be an arbitrary policy, and $t \geq 1$. Then*

$$\mathbb{E}F^\pi(t) \leq \mathbb{E}F^*(t).$$

The optimality of OCL crucially relies on assumption (i). Consider for example the following problem instance: $\mathcal{X} = \{1,2,3,4\}$, $A = \{1,2,3\}$, $K = 3$, $\nu_1 = \delta_1$, $\nu_2 = \frac{2}{5}(\delta_1 + \delta_2) + \frac{1}{5}\delta_4$, and $\nu_3 = \frac{2}{5}(\delta_1 + \delta_3) + \frac{1}{5}\delta_4$ and $t = 2$. In this case OCL first chooses expert 1, and then (say) expert 2: this yields $F^*(2) = 1 + 2/5 = 7/5$. But the strategy π consisting in choosing first expert 2, and then expert 3, is readily seen to have expected return $\mathbb{E}F^\pi(2) = 2/5 \times (1 + 2/5) + 2/5 \times (1 + 4/5) + 1/5 \times 4/5 = 36/25 > 7/5$.

The next lemma is a technical result on OCL that shall prove to be very useful to derive a standard regret bound for Good-UCB. Its proof is also given in the appendix.

Lemma 3 *Let π be an arbitrary policy, and for $t \geq 1$ let*

$$\bar{I}_t = \arg\max_{1 \leq i \leq K} R_{i, n_{i,t-1}^\pi} .$$

Then

$$\mathbb{E}F^*(t) \leq \sum_{s=1}^t \mathbb{E}R_{\bar{I}_s, n_{\bar{I}_s, s-1}^\pi}.$$

2.4 Classical Analysis of the Good-UCB Algorithm

We provide here an upper bound on the expectation of $F^*(t) - F^{UCB}(t)$ which is completely distribution-free, and which depends only on the horizon t and on the number K of experts. This bound grows like $O(\sqrt{Kt\log(t)})$, which is a usual rate for a bandit problem. Indeed, thanks to Lemma 3, the analysis presented in this section follows the lines of classical regret analyses, see for instance Bubeck and Cesa-Bianchi (2012, and the references therein). Below, we discuss some differences between the discovery problem considered here and bandit problems, and we provide an alternative analysis of the Good-UCB algorithm which is more suited to understand its long-term behavior.

Theorem 4 *For any $t \geq 1$, under assumption (i), Good-UCB (with constant $C = (1+\sqrt{2})\sqrt{3}$) satisfies*

$$\mathbb{E}\left[F^*(t) - F^{UCB}(t)\right] \leq 17\sqrt{Kt\log(t)} + 20\sqrt{Kt} + K + K\log(t/K) .$$

Proof Consider the event

$$\xi = \Bigg\{ \forall i \in \{1,\ldots,K\}, \forall u > \sqrt{Kt}, \forall s \leq u, \\ \hat{R}_{i,s} - \frac{1}{s} - (1+\sqrt{2})\sqrt{\frac{3\log(4u)}{s}} \leq R_{i,s} \leq \hat{R}_{i,s} + (1+\sqrt{2})\sqrt{\frac{3\log(4u)}{s}} \Bigg\} .$$

Using Proposition 1 and an union bound, one obtains $\mathbb{P}(\xi) \geq 1 - \sqrt{\frac{K}{t}}$, and thus

$$\mathbb{E}\left[(F^*(t) - F^{UCB}(t))(1 - \mathbb{1}_\xi)\right] \leq t\sqrt{\frac{K}{t}} = \sqrt{Kt}\,.$$

Let $u > \sqrt{Kt}$ and $\bar{I}_u = \arg\max_{1\leq i\leq K} R_{i,n_{i,u-1}}$ be defined as in Lemma 3. On the event ξ, one obtains by definition of I_u that

$$\begin{aligned}
R_{I_u,n_{I_u,u-1}} &\geq \hat{R}_{I_u,n_{I_u,u-1}} - \frac{1}{n_{I_u,u-1}} - (1+\sqrt{2})\sqrt{\frac{3\log(4u)}{n_{I_u,u-1}}} \\
&\geq \hat{R}_{I_u,n_{I_u,u-1}} + (1+\sqrt{2})\sqrt{\frac{3\log(4u)}{n_{I_u,u-1}}} - \frac{1}{n_{I_u,u-1}} - 2(1+\sqrt{2})\sqrt{\frac{3\log(4u)}{n_{I_u,u-1}}} \\
&\geq \hat{R}_{\bar{I}_u,n_{\bar{I}_u,u-1}} + (1+\sqrt{2})\sqrt{\frac{3\log(4u)}{n_{I^*_u,u-1}}} - \frac{1}{n_{I_u,u-1}} - 2(1+\sqrt{2})\sqrt{\frac{3\log(4u)}{n_{I_u,u-1}}} \\
&\geq R_{\bar{I}_u,n_{\bar{I}_u,u-1}} - \frac{1}{n_{I_u,u-1}} - 2(1+\sqrt{2})\sqrt{\frac{3\log(4u)}{n_{I_u,u-1}}}\,,
\end{aligned}$$

and thus

$$\begin{aligned}
R_{\bar{I}_u,n_{\bar{I}_u,u-1}} - R_{I_u,n_{I_u,u-1}} &\leq \frac{1}{n_{I_u,u-1}} + 2(1+\sqrt{2})\sqrt{\frac{3\log(4u)}{n_{I_u,u-1}}} \\
&\leq \frac{1}{n_{I_u,u-1}} + 2(1+\sqrt{2})\sqrt{\frac{3\log(4t)}{n_{I_u,u-1}}}\,.
\end{aligned}$$

Hence, using Lemma 3 and the above computation, one obtains

$$\begin{aligned}
\mathbb{E}\left[F^*(t) - F^{UCB}(t)\right] &\leq \sqrt{Kt} + \mathbb{E}\left[\sum_{u=1}^{t} \frac{1}{n_{I_u,u-1}} + 2(1+\sqrt{2})\sqrt{\frac{3\log(4t)}{n_{I_u,u-1}}}\right] \\
&= \sqrt{Kt} + \mathbb{E}\left[\sum_{i=1}^{K}\sum_{s=1}^{n_{i,t-1}} \frac{1}{s} + 2(1+\sqrt{2})\sqrt{\frac{3\log(4t)}{s}}\right] \\
&\leq \sqrt{Kt} + \mathbb{E}\left[\sum_{i=1}^{K} 1 + \log(n_{i,t-1}) + 4(1+\sqrt{2})\sqrt{3\log(4t)(n_{i,t-1}+1)}\right] \\
&\leq \sqrt{Kt} + K + K\log(t/K) + 4(1+\sqrt{2})\sqrt{3Kt\log(4t)}
\end{aligned}$$

by Jensen's inequality and the fact that $\sum_{i=1}^{K} n_{i,t-1} = t-1$. ■

The cumulative regret bound provided in Theorem 4 has a similar flavor as well known regret bounds for the multi-armed bandit problem. Unfortunately here, such bounds, by suggesting that the regret increases with t, do not represent completely the behavior of Good-UCB: as we shall see

in the experiments, the difference between $F^*(t)$ and $F^{UCB}(t)$ is bounded and tends to 0 as t tends to infinity (indeed, ultimately any reasonable strategy will find all the interesting items). Theorem 4 provides insight into the properties of Good-UCB only for 'small' values of t.

The weakness of Theorem 4 and its analysis is that, by using the upper bound of Lemma 3, one ignores the *restoring property* of the game: if a policy makes a sub-optimal choice at some given time t, then it will have better opportunities than OCL in the future, which prevents the regret from growing too much. In the next section we provide a completely different analysis of Good-UCB that takes advantage of this restoring property. This results in a non-standard regret bound, which differs from usual results in the multi-armed bandit literature.

Let us make one more comment about the bound of Theorem 4. On the contrary to the multi-armed bandit, the discovery problem discussed in this paper has a 'natural' time scale: if the horizon t is too small, then even OCL will not be able to discover a significant proportion of interesting items, while if t is too large then any reasonable strategy will find almost all interesting items. To go around this issue we find it more elegant to study the waiting time $T(\lambda)$ (see (2)) which yields a sort of automatic normalization of the time scale.

2.5 Time-uniform Analysis of the Good-UCB Algorithm

In this section we analyze the waiting time of Good-UCB under assumption (i). We shall derive a non-linear regret bound as follows. For a fixed $\lambda \in (0,1)$ we consider the number of requests $T_{UCB}(\lambda)$ that Good-UCB needs to make in order to have a missing mass of interesting items smaller than λ on each expert, see (2). We also consider the omniscient oracle strategy that minimizes this number of requests, given the knowledge of λ and the sequence of answers to the requests $(X_{i,s})_{1\le i\le K, s\ge 1}$. We denote by $T^*(\lambda)$ the corresponding number of requests for this omniscient oracle strategy. (Note that this strategy is even more powerful than the OCL studied in the previous sections.) We now prove that with high probability, $T_{UCB}(\lambda)$ is smaller than $T^*(\lambda')$, for some λ' close to λ and up to a small additional term.

Theorem 5 *Let $c > 0$ and $S \ge 1$. Under assumption (i), Good-UCB (with constant $C = (1+\sqrt{2})\sqrt{c+2}$) satisfies with probability at least $1 - \frac{K}{cS^c}$, for any $\lambda \in (0,1)$,*

$$T_{UCB}(\lambda) \le T^* + KS\log\left(8T^* + 16KS\log(KS)\right),$$

$$\text{where} \quad T^* = T^*\left(\lambda - \frac{3}{S} - 2(1+\sqrt{2})\sqrt{\frac{c+2}{S}}\right).$$

Informally this bound shows that Good-UCB slightly lags behind the omniscient oracle strategy. Under more restrictive assumptions on the experts it is possible to obtain a more explicit bound by studying the variations of T. In the next section we take another route and we show that the above upper bound can be used to prove a clear qualitative property for Good-UCB, namely its *macroscopic optimality*.

Proof Recall that we work under assumption (i), and we run Good-UCB with parameter $C = (1+\sqrt{2})\sqrt{c+2}$, for some positive constant c. After t pulls, the missing mass estimate of expert i is:

$$\hat{R}_{i,t} = \frac{1}{t}\sum_{x\in A} \mathbb{1}\left\{1 = \sum_{s=1}^{t} \mathbb{1}\{X_{i,s} = x\}\right\}.$$

We consider the following event:

$$\xi = \Bigg\{ \forall i \in \{1,\dots,K\}, \forall t > S, \forall s \leq t, \\ \hat{R}_{i,s} - \frac{1}{s} - (1+\sqrt{2})\sqrt{\frac{(c+2)\log(4t)}{s}} \leq R_{i,s} \leq \hat{R}_{i,s} + (1+\sqrt{2})\sqrt{\frac{(c+2)\log(4t)}{s}} \Bigg\}.$$

Using Proposition 1 and an union bound, one obtains $\mathbb{P}(\xi) \geq 1 - \frac{K}{cS^c}$. In the following we work on the event ξ. Recall that $T^*(\lambda)$ (respectively $T_{UCB}(\lambda)$) is the time at which the omniscient oracle strategy (respectively the Good-UCB strategy) attains a missing mass smaller than λ on all experts. Note that $T^*(\lambda)$ and $T_{UCB}(\lambda)$ are functions of $(X_{i,s})_{1\leq i\leq K, s\geq 1}$. In particular one can write:

$$T_{UCB}(\lambda) = \min\left\{ t \geq 1 : \forall i \in \{1,\dots,K\}, R_{i,n_{i,t}} \leq \lambda \right\},$$
$$T^*(\lambda) = \sum_{i=1}^{K} T_i^*(\lambda), \text{ where } T_i^*(\lambda) = \min\left\{ t \geq 1 : R_{i,t} \leq \lambda \right\}.$$

Let

$$U(\lambda) = \min\left\{ t \geq 1 : \forall i \in \{1,\dots,K\}, \hat{R}_{i,n_{i,t}} + (1+\sqrt{2})\sqrt{\frac{(c+2)\log(4t)}{n_{i,t}}} \leq \lambda \right\}.$$

Let $S' \geq S$ to be defined later. On the event ξ one clearly gets $T_{UCB}(\lambda) \leq \max(S', U(\lambda))$. Moreover the following inequalities hold true if $U(\lambda) > S'$ (see below for an explanation of each inequality)

$$\begin{aligned} R_{i,n_{i,U(\lambda)}} &\geq \hat{R}_{i,n_{i,U(\lambda)}} - \frac{1}{n_{i,U(\lambda)}} - (1+\sqrt{2})\sqrt{\frac{(c+2)\log(4U(\lambda))}{n_{i,U(\lambda)}}} \\ &\geq \hat{R}_{i,n_{i,U(\lambda)}-1} - \frac{3}{n_{i,U(\lambda)}} - (1+\sqrt{2})\sqrt{\frac{(c+2)\log(4U(\lambda))}{n_{i,U(\lambda)}}} \\ &\geq \left(\lambda - (1+\sqrt{2})\sqrt{\frac{(c+2)\log(4U(\lambda))}{n_{i,U(\lambda)}-1}}\right) - \frac{3}{n_{i,U(\lambda)}} - (1+\sqrt{2})\sqrt{\frac{(c+2)\log(4U(\lambda))}{n_{i,U(\lambda)}}} \\ &\geq \lambda - \frac{3}{n_{i,U(\lambda)}} - 2(1+\sqrt{2})\sqrt{\frac{(c+2)\log(4U(\lambda))}{n_{i,U(\lambda)}-1}}. \end{aligned}$$

The first inequality comes from the fact that we are on event ξ and we assume $U(\lambda) > S'$. The second inequality uses the fact that when we make a request to an expert, the number of items uniquely seen on this expert can drop by at most one, and thus we get

$$s\hat{R}_{i,s} \geq (s-1)\hat{R}_{i,s-1} - 1 \geq s\hat{R}_{i,s-1} - 2.$$

The third inequality is the key step of the proof. Consider the time step t such that $n_{i,t} = n_{i,U(\lambda)} - 1$ and $n_{i,t+1} = n_{i,U(\lambda)}$. Since $t < U(\lambda)$ we know that one of the expert satisfies $\hat{R}_{j,n_{j,t}} + (1+\sqrt{2})\sqrt{\frac{(c+2)\log(4t)}{n_{j,t}}} > \lambda$. Moreover, since Good-UCB is run with constant $C = (1+\sqrt{2})\sqrt{c+2}$ and since we make a request to expert i at time t, we know that it maximizes the Good-UCB index, and thus $\hat{R}_{i,n_{i,t}} + (1+\sqrt{2})\sqrt{\frac{(c+2)\log(4t)}{n_{i,t}}} > \lambda$. Using that $t \leq U(\lambda)$ completes the proof of the third inequality. The fourth inequality is trivial.

We just proved that if $n_{i,U(\lambda)} > S'$ then

$$R_{i,n_{i,U(\lambda)}} \geq \lambda - \frac{3}{S'} - 2(1+\sqrt{2})\sqrt{\frac{(c+2)\log(4U(\lambda))}{S'}},$$

which clearly implies

$$n_{i,U(\lambda)} \leq T_i^* \left(\lambda - \frac{3}{S'} - 2(1+\sqrt{2})\sqrt{\frac{(c+2)\log(4U(\lambda))}{S'}} \right).$$

Thus in any case we have proved that

$$n_{i,U(\lambda)} \leq S' + T_i^* \left(\lambda - \frac{3}{S'} - 2(1+\sqrt{2})\sqrt{\frac{(c+2)\log(4U(\lambda))}{S'}} \right),$$

which implies

$$\begin{aligned} U(\lambda) &\leq KS' + T^* \left(\lambda - \frac{3}{S'} - 2(1+\sqrt{2})\sqrt{\frac{(c+2)\log(4U(\lambda))}{S'}} \right) \\ &\leq KS\log(4U(\lambda)) + T^* \left(\lambda - \frac{3}{S} - 2(1+\sqrt{2})\sqrt{\frac{c+2}{S}} \right), \end{aligned}$$

where the last inequality follows by taking $S' = S\log(4U(\lambda))$. Finally using Lemma 9 (in the appendix) and $T_{UCB}(\lambda) \leq \max(S', U(\lambda))$ ends the proof. ■

3. Macroscopic Limit

In the previous section we derived a very general non-linear regret bound for Good-UCB. Here we shall study the behavior of Good-UCB under more restrictive assumptions on the experts, but it will allow us to derive a clear qualitative statement about its performance, and it also permits easier comparison with other strategies such as uniform sampling. In this section we shall add the two following assumptions in addition to assumption (i):

(ii) finite supports with the same cardinality: $|\mathrm{supp}(P_i)| = N, \forall i \in \{1,\dots,K\}$,

(iii) uniform distributions: $P_i(x) = \frac{1}{N}, \forall x \in \mathrm{supp}(P_i), \forall i \in \{1,\dots,K\}$.

These assumptions are primarily made in order to be able to assess the performance of the optimal strategy. In this setting it is convenient to re-parameterize slightly the problem (in particular we make explicit the dependency on N for reasons that will appear later). Let $\mathcal{X}^N = \{1,\dots,K\} \times \{1,\dots,N\}$, $A^N \subset \mathcal{X}^N$ the set of interesting items of $\mathcal{X}^N$, and $Q^N = |A^N|$ the number of interesting items. We assume that, for expert $i \in \{1,\dots,K\}$, P_i^N is the uniform distribution on $\{i\} \times \{1,\dots,N\}$. We also denote by $Q_i^N = \left|A^N \cap (\{i\} \times \{1,\dots,N\})\right|$ the number of interesting items accessible through requests to expert i. Without loss of generality, we assume in this section that $Q_1^N \geq Q_2^N \geq \cdots \geq Q_K^N$.

The macroscopic limit that we investigate in this section corresponds to the setting where N goes to infinity together with the Q_i^N in such a way that $Q_i^N/N \to q_i \in (0,1)$. For a given strategy we are interested in the time $T^N(\lambda)$ such that all experts have at most $N\lambda$ undiscovered interesting items. In particular we define $T^N_{UCB}(\lambda)$ (respectively $T^N_*(\lambda)$) to be the corresponding time for the Good-UCB strategy (respectively the oracle omniscient strategy). In the macroscopic limit we shall be particularly interested in normalized limit waiting time $\lim_{N\to+\infty} T^N(\lambda)/N$.

3.1 Macroscopic Behavior of the Oracle Closed-loop Strategy

In this section we shall derive an explicit upper bound on the macroscopic limit of T^N_* by studying the OCL strategy introduced in Section 2.3. Recall that at each time step, OCL makes a request to one of the experts with highest number of still undiscovered interesting items: the expert requested at time t is:

$$I_t \in \operatorname*{arg\,max}_{1\le i\le K} P_i\left(A \setminus \{X_{1,1},\dots,X_{1,n_{1,t}},\dots,X_{K,1},\dots,X_{K,n_{K,t}}\}\right) .$$

Theorem 6 *For every $\lambda \in (0,q_1)$, for every sequence $(\lambda^N)_N$ converging to λ as N goes to infinity, under assumption (i), (ii) and (iii), almost surely*

$$\lim_{N\to\infty} \frac{T^N_{OCL}(\lambda^N)}{N} = \sum_{i:q_i>\lambda} \log\frac{q_i}{\lambda} .$$

Proof Denote by B_i^N the set of interesting items in $\{1,\dots,N\}$ supported by P_i^N: $B_i^N = \{x \in \{1,\dots,N\} : (i,x) \in A^N\}$. Successive draws of expert i are denoted $(i,X_{i,1}^N),(i,X_{i,2}^N)\dots$ where the variables $(X_{i,n}^N)_{i,n}$ are assumed to be independent. Without loss of generality, we may assume that $N\lambda^N$ is a positive integer, for otherwise λ^N can be replaced by $\lceil N\lambda^N\rceil/N$. We denote by $(D_{i,k}^N)_{1\le k\le Q_i^N}$ the increasing sequence of the indices corresponding to draws for which new interesting items are discovered with expert i:

$$D_{i,1}^N = \min\left\{n \ge 1 : X_{i,n}^N \in B_i^N\right\}, \quad D_{i,2}^N = \min\left\{n \ge D_{i,1}^N : X_{i,n}^N \in B_i^N \setminus \left\{X_{i,D_{i,1}^N}^N\right\}\right\},\dots$$

We also define $S_{i,0}^N = 0$ and for $k \ge 1, S_{i,k}^N = D_{i,k}^N - D_{i,k-1}^N$. The random variables $S_{i,k}^N$ $(1 \le i \le K, k \ge 1)$ are independent with geometric distribution $\mathcal{G}((1+Q_i^N-k)/N)$.

At every step, the OCL should call the expert with maximal number of undiscovered interesting items. Hence, it can:

- first request expert 1 for $D^N_{1,Q_1^N-Q_2^N}$ steps;
- then, alternatively request
  - expert 1 for $S^N_{1,1+Q_1^N-Q_2^N}$ steps;
  - expert 2 for $S^N_{2,1}$ steps;
  - expert 1 for $S^N_{1,2+Q_1^N-Q_2^N}$ steps;
  - expert 2 for $S^N_{2,2}$ steps;
  - and so on, until there are only Q_3^N undiscovered interesting items on experts 1 and 2.

- and so on, including successively experts $3, 4, \ldots, K$ in the alternation.

Obviously,

$$T^N_{OCL}(\lambda^N) = \sum_{i:Q^N_i > N\lambda^N} D^N_{i,Q^N_i - N\lambda^N} \cdot$$

It suffices now to show that for every expert $i \in \{1, \ldots, K\}$, $D^N_{i,Q^N_i - N\lambda^N}/N$ converges almost surely to $\log(q_i/\lambda)$ as N goes to infinity. Write

$$W^N_{i,N\lambda^N} = \frac{1}{N}\left(D^N_{i,Q^N_i - N\lambda^N} - \mathbb{E}\left[D^N_{i,Q^N_i - N\lambda^N}\right]\right) = \frac{1}{N}\sum_{k=1}^{Q^N_i - N\lambda^N - 1}\left(S^N_{i,k} - \mathbb{E}\left[S^N_{i,k}\right]\right) . \tag{5}$$

For every positive integer d and for $k \in \{1, \ldots, N\lambda^N - 1\}$, elementary manipulations of the geometric distribution yield that

$$\mathbb{E}\left[\left(S^N_{i,k} - \mathbb{E}\left[S^N_{i,k}\right]\right)^d\right] \leq \mathbb{E}\left[\left(S^N_{i,N\lambda^N} - \mathbb{E}\left[S^N_{i,N\lambda^N}\right]\right)^d\right] \leq \frac{c(d)}{(\lambda^N)^d} \leq \frac{2c(d)}{\lambda^4}$$

for some positive constant $c(d)$ depending only on d, and for N large enough. Hence, taking (5) to the fourth power and developing yields

$$\mathbb{E}\left[\left(W^N_{i,N\lambda^N}\right)^4\right] \leq \frac{c'}{N^2\lambda^4}$$

for some positive constant c'. Using Markov's inequality together with the Borel-Cantelli lemma, this permits to show that W^N_{i,λ^N} converges almost surely to 0 as N goes to infinity. But

$$\frac{1}{N}\mathbb{E}\left[D^N_{i,Q^N_i - N\lambda^N}\right] = \frac{1}{Q^N_1} + \cdots + \frac{1}{N\lambda^N + 1} = \log\frac{Q^N_i}{N\lambda^N} - \varepsilon^N ,$$

with $0 \leq \varepsilon^N \leq 1/(N\lambda^N)$ according to Lemma 10, and thus

$$\frac{1}{N}\mathbb{E}\left[D^N_{i,Q^N_i - N\lambda^N}\right] \to \lim_{N\to\infty} \log\left(\frac{Q^N_i/N}{\lambda^N}\right) = \log(q_i/\lambda) ,$$

which concludes the proof. ■

3.2 Macroscopic Behavior of Uniform Sampling

In this section we study the simple uniform sampling strategy that cycles through the experts, that is, at time t uniform sampling makes a request to the $(t \mod [K])^{th}$ expert. This strategy is not macroscopically optimal unless all experts have the same number of interesting items. Furthermore the next proposition makes precise the extent of improvement of a macroscopic optimal strategy over uniform sampling. The proof follows the exact same steps than the proof of Theorem 6 and thus is omitted.

Proposition 7 *For every $\lambda \in (0, q_1)$, for every sequence $(\lambda^N)_N$ converging to λ as N goes to infinity, under assumption (i), (ii) and (iii), almost surely*

$$\lim_{N\to\infty} \frac{T^N_{US}(\lambda^N)}{N} = K \log\frac{q_1}{\lambda} .$$

3.3 Macroscopic Optimality of Good-UCB

Using the regret bound of Theorem 5 we obtain the following corollary that shows the asymptotic optimality of the Good-UCB algorithm in the macroscopic sense.

Corollary 8 *Take $C = (1+\sqrt{2})\sqrt{c+2}$ with $c > 3/2$ in Algorithm 1. Under assumption (i), (ii) and (iii), for every sequence $(\lambda^N)_N$ converging to λ as N goes to infinity, almost surely*

$$\limsup_{N\to+\infty} \frac{T^N_{UCB}(\lambda^N)}{N} \leq \sum_{i:q_i>\lambda} \log\frac{q_i}{\lambda}\,.$$

Proof Let $S^N = N^{2/3}$. First note that:

$$\ell^N \overset{def}{=} \lambda^N - \frac{3}{S^N} - 2(1+\sqrt{2})\sqrt{\frac{c+2}{S^N}} \to \lambda \quad \text{when } N\to\infty\,.$$

Thus, by Theorem 6, and the fact that the OCL strategy needs at least as much time as the omniscient oracle strategy in order to find the same number of items, there exists an event Ω of probability 1 on which

$$\limsup_{N\to+\infty} \frac{T^N_*\left(\ell^N\right)}{N} \leq \sum_{i:q_i>\lambda} \log\frac{q_i}{\lambda}\,.$$

Thus, according to Theorem 5, for each positive integer N there exists an event A_N of probability $P(A_N) \geq 1 - K/(cN^{2c/3})$ on which

$$\begin{aligned}\frac{T^N_{UCB}(\lambda^N)}{N} &\leq \frac{T^N_*\left(\ell^N\right)}{N} + \frac{KS^N}{N}\log\left(8T^N_*\left(\ell^N\right) + 16KS\log(KS^N)\right)\\ &= \frac{T^*_N\left(\ell^N\right)}{N} + O\left(\frac{\log(N)}{N^{1/3}}\right)\,.\end{aligned}$$

Using Borel-Cantelli's lemma and the fact that, with our choice of parameters, $\sum_N N^{-2c/3} < \infty$, we obtain that except maybe on the set (of probability 0) $\bar{\Omega}\cup\limsup\overline{A_N}$,

$$\limsup_{N\to\infty} \frac{T^N_{UCB}(\lambda^N)}{N} \leq \limsup_{N\to+\infty} \frac{T^N_*(\ell^N)}{N} \leq \sum_{i:q_i>\lambda} \log\frac{q_i}{\lambda}\,,$$

which ends the proof. ■

4. Simulations

We provide a few simulations illustrating the behavior of the Good-UCB algorithm and the asymptotic analysis above of Section 3. We first consider an example with $K = 7$ different sampling distributions satisfying assumptions [(i),(ii),(iii)], with respective proportions of interesting items $q_1 = 51.2\%, q_2 = 25.6\%, q_3 = 12.8\%, q_4 = 6.4\%, q_5 = 3.2\%, q_6 = 1.6\%$ and $q_7 = 0.8\%$.

We have chosen to display here the numbers of items found as a function of the number of draws (see (1)), instead of the times $T^N(\lambda^N)$, because they express more intuitively the discovering

possibilities of each algorithm. Note, however, that the correspondence between these two quantities is straightforward, especially in the macroscopic limit: For $\lambda \in (0, q_1)$ let

$$T(\lambda) = \sum_{i:q_i>\lambda} \log \frac{q_i}{\lambda} \,. \tag{6}$$

It is easy to show that the proportion of interesting items found by the OCL strategy after Nt draws converge to

$$F(t) = \sum_{i=1}^{K} \left(q_i - T^{-1}(t)\right)_+ \,. \tag{7}$$

Furthermore the latter expression is a lower bound for the corresponding proportion of interesting items found by the Good-UCB algorithm. Proposition 11, proved in the Appendix, provides a more explicit expression for F: denoting $q = \sum_{i=1}^{K} q_i$, there exists an increasing, $\{1,\dots,K\}$-valued function I such that, for each t,

$$F(t) = q - I(t)\underline{q}_{I(t)} \exp\left(-t/I(t)\right) \,,$$

where $\underline{q}_{I(t)}$ denotes the geometric mean of $q_1,\dots,q_{I(t)}$. This permits an explicit comparison of the macroscopic performance of the Good-UCB algorithm with uniform sampling: when all distributions are sampled equally often, the proportion of unseen interesting items at time t is smaller than

$$\sum_{i=1}^{K} q_i \exp(-t/K) = K\bar{q}_K \exp(-t/K) \,,$$

where $\bar{q}_K = (\sum_{i=1}^{K} q_i)/K$ is the arithmetic mean of the $(q_i)_i$. On the other hand, for the Good-UCB algorithm, the proportion of unseen interesting items at time t is smaller than

$$I(t)\underline{q}_{I(t)} \exp\left(-t/I(t)\right) \,.$$

The ratio of those two quantities is a decreasing function of time lower-bounded by $\bar{q}_K/\underline{q}_K \geq 1$, the ratio of the arithmetic mean with the geometric mean of the $(q_i)_i$. As expected, this ratio gets larger when the proportions of interesting items among experts becomes more unbalanced.

Figure 1 displays the number of items found as a function of time by the Good-UCB (solid), the OCL (dashed) and the uniform sampling scheme that alternates between experts (dotted). The results are presented for sizes $N = 128, N = 500, N = 1000$ and $N = 10000$, each time for one representative run (averaging over different runs removes the interesting variability of the process). We chose to plot the number of items found rather than the waiting time t as the former is easier to visualize while the latter was easier to analyze. In fact, macroscopic optimality in terms of number of items found could also be derived with the techniques of Section 3. Figure 1 also shows clearly the macroscopic convergence of Good-UCB to the OCL. Moreover, it can be seen that, even for very moderate values of N, the Good-UCB significantly outperforms uniform sampling even if it is clearly distanced by the OCL.

For these simulations, the parameter C of Algorithm Good-UCB has been taken equal to $1/2$, which is a rather conservative choice. In fact, it appears that during all rounds of all runs, all upper-confidence bounds did contain the actual missing mass. Of course, a bolder choice of C can only improve the performance of the algorithm, as long as the confidence level remains sufficient.

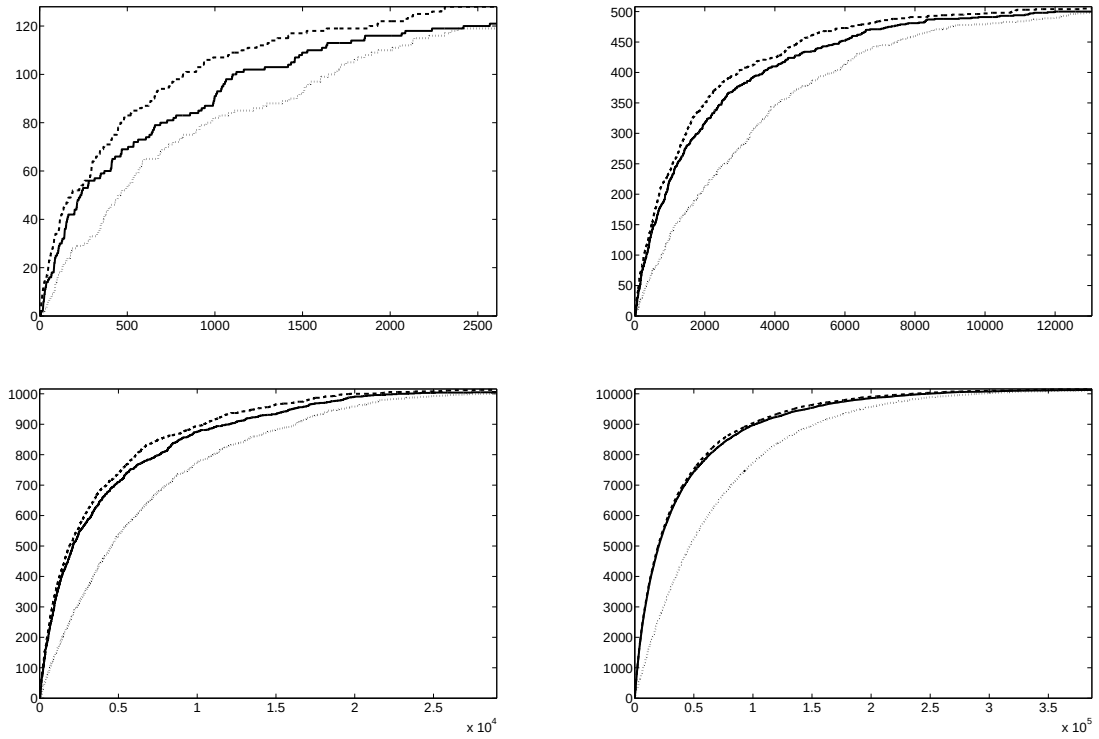

Figure 1: Number of items found by Good-UCB (solid), the OCL (dashed), and uniform sampling (dotted) as a function of time for sizes $N = 128, N = 500, N = 1000$ and $N = 10000$ in a 7-experts setting.

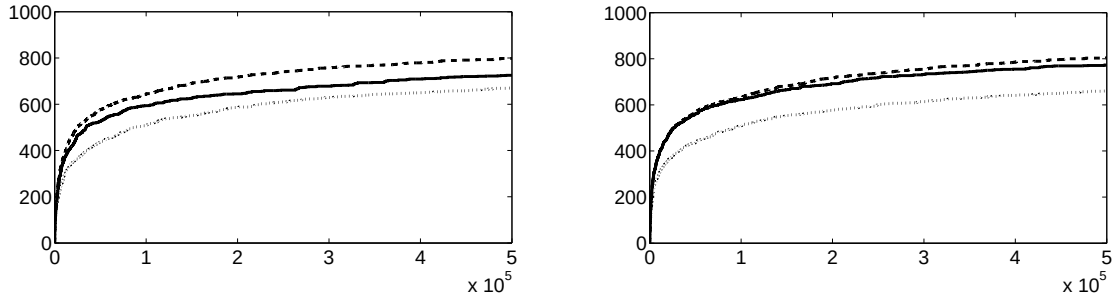

Figure 2: Number of prime numbers found by Good-UCB (solid), the OCL (dashed), and uniform sampling (dotted) as a function of time, using geometric experts with means $100, 300, 500, 700$ and 900, for $C = 0.1$ (left) and $C = 0.02$ (right).

In order to illustrate the efficiency of the Good-UCB algorithm in a more difficult setting, which does not satisfy any of the assumptions (i), (ii) and (iii), we also considered the following (artificial) example: $K = 5$ probabilistic experts draw independent sequences of geometrically distributed

random variables, with expectations 100, 300, 500, 700 and 900 respectively. The set of interesting items is the set of prime numbers. We compare the oracle closed-loop policy, Good-UCB and uniform sampling. The results are displayed in Figure 2. Even if the difference remains significant between Good-UCB and the OCL, the former still performs significantly better than uniform sampling during the entire discovery process. In this example, choosing a smaller parameter C seems to be preferable; this is due to the fact that the proportion of interesting items on each arm is low; in that case, it may be possible to show, by using tighter concentration inequalities, that the concentration of the Good-Turing estimator is actually better than suggested by Proposition 1. In fact, this experiment suggests that the value of C should be chosen smaller when the remaining missing mass is small.

Acknowledgments

We are especially thankful to one of the anonymous referees for suggesting to us to write Sections 2.3 and 2.4.

Appendix A.

Proof of lemma 2 We proceed by induction on t. For $t = 1$, the result is obvious. For $t > 1$, denote by $\bar{\pi}$ the policy choosing $I_1^{\bar{\pi}} = I_1^{\pi}$ and then playing like OCL for the $t-1$ remaining rounds. Denote $H_1 = (I_1^{\pi}, X_{I_1^{\pi},1})$, and $F^{\pi}(2:t)$ (respectively $F^{\bar{\pi}}(2:t)$) the number of interesting items found by policy π (respectively $\bar{\pi}$) between rounds 2 and t. Note that conditionally on H_1, $F^{\bar{\pi}}(2:t)$ corresponds to $F^*(t-1)$ in some modified problem (where one interesting item on expert I_1^{π} might have been removed from the set of interesting items). Thus one can apply the induction hypothesis to obtain

$$\mathbb{E}\left[F^{\pi}(2:t)|H_1\right] \leq \mathbb{E}\left[F^{\bar{\pi}}(2:t)|H_1\right].$$

Let us assume in the following that I_1^{π} is deterministic (we make this assumption only for sake of clarity, everything go through with a randomized choice of I_1^{π}). Then thanks to the above inequality one has

$$\mathbb{E}F^{\pi}(t) = R_{I_1^{\pi},0} + \mathbb{E}\left[F^{\pi}(2:t)\right] \leq R_{I_1^{\pi},0} + \mathbb{E}\left[F^{\bar{\pi}}(2:t)\right] = \mathbb{E}F^{\bar{\pi}}(t). \tag{8}$$

Now let

$$\tau = \min\{s \geq 1 : I_s^* = I_1^{\pi}\}.$$

On the event $\tau \leq t$, OCL and $\bar{\pi}$ observe exactly the same items during the t first rounds, and thus

$$F^{\bar{\pi}}(t)\mathbb{1}\{\tau \leq t\} = F^*(t)\mathbb{1}\{\tau \leq t\}. \tag{9}$$

On the other hand on the event $\tau > t$, $\bar{\pi}$ observe the same items between rounds 2 and t than OCL between rounds 1 and $t-1$, that is $F^{\bar{\pi}}(2:t)\mathbb{1}\{\tau > t\} = F^*(t-1)\mathbb{1}\{\tau > t\}$. Thanks to assumption (i), this implies (denoting $Y_1^*, \ldots, Y_t^*$ for the sequence of items observed by OCL),

$$F^{\bar{\pi}}(t)\mathbb{1}\{\tau > t\} = \left(\mathbb{1}\{X_{I_1^{\pi},1} \in A\} + F^*(t) - \mathbb{1}\{Y_t^* \in A \setminus \{Y_1^*, \ldots, Y_{t-1}^*\}\}\right)\mathbb{1}\{\tau > t\}. \tag{10}$$

By combining (8), (9) and (10), it only remains to show that

$$\mathbb{E}[\mathbb{1}\{X_{I_1^{\pi},1} \in A\}\mathbb{1}\{\tau > t\}] \leq \mathbb{E}[\mathbb{1}\{Y_t^* \in A \setminus \{Y_1^*, \ldots, Y_{t-1}^*\}\}\mathbb{1}\{\tau > t\}]. \tag{11}$$

Since $X_{I_1^\pi,1}$ is independent of $\mathbb{1}\{\tau > t\}$, one has $\mathbb{E}[\mathbb{1}\{X_{I_1^\pi,1} \in A\}\mathbb{1}\{\tau > t\}] = \mathbb{E}[R_{I_1^\pi,0}\mathbb{1}\{\tau > t\}]$. Moreover, noting that $\mathbb{1}\{\tau > t\}$, I_t^* and $R_{I_t^*,n^*_{I_t^*,t-1}}$ are measurable with respect to $H^*_{t-1} = \left(I_1^*, Y_1^*, \ldots, I_{t-1}^*, Y_{t-1}^*\right)$, one has

$$\mathbb{E}[\mathbb{1}\{Y_t^* \in A \setminus \{Y_1^*, \ldots, Y_{t-1}^*\}\}\mathbb{1}\{\tau > t\}] = \mathbb{E}[R_{I_t^*,n^*_{I_t^*,t-1}}\mathbb{1}\{\tau > t\}].$$

Finally remark that on the event $\tau > t$ one necessarily have that the remaining missing mass on the expert pulled at time t by OCL is larger than the initial missing mass of expert I_1^π, that is $R_{I_t^*,n^*_{I_t^*,t-1}}\mathbb{1}\{\tau > t\} \geq R_{I_1^\pi,0}\mathbb{1}\{\tau > t\}$, which concludes the proof of (11). ■

Proof of Lemma 3 Let $Y_s^\pi = X_{I_s^\pi,n_{I_s^\pi,s}}$ be the item observed by π at time step s, and $H_s^\pi = \left(I_1^\pi, Y_1^\pi, \ldots, I_{s-1}^\pi, Y_{s-1}^\pi\right)$ be the history of π prior to making the decision on time s. For any history $h_s = (i_1, y_1, \ldots, i_{s-1}, y_{s-1})$, let $F^*(t|h_s)$ be the number of newly discovered interesting items when running OCL from the history h_s for $t-s+1$ steps. 'From the history h_s' means that, prior to running OCL, the sequence of experts $i_1, \ldots, i_{s-1}$ has been chosen and has led to the observations $y_1, \ldots, y_{s-1}$. For $s' \geq s$ we shall also denote $I_{s'}^*(h_s)$ (respectively $Y_{s'}^*(h_s)$) the sequence of expert requests made by OCL starting at h_s (respectively the corresponding sequence of observed items). Note in particular that $\bar{I}_s$ defined in the statement of the lemma corresponds to $I_s^*(H_s^\pi)$. We shall also need τ_s to be the first time when OCL, running from history H_s^π, selects expert I_s^π, that is

$$\tau_s = \min\{s' \geq s : I_{s'}^*(H_s^\pi) = I_s^\pi\}\,,$$

and $\tau_s = +\infty$ if there is no interesting item to be found by expert I_s^π at time s.

We shall prove that

$$\mathbb{E}[F^*(t|H_s^\pi) - F^*(t|H_{s+1}^\pi)] \leq \mathbb{E}R_{\bar{I}_s,n^\pi_{\bar{I}_s,s-1}}, \tag{12}$$

which inductively yields the lemma since $F^*(t) = F^*(t|h_1)$ and $F^*(t|h_{t+1}) = 0$.

First let us consider the case when $\tau_s \leq t$. Then the observed items with OCL (running from H_s^π) between step s and t remains unchanged if one forces OCL to play I_s^π at time step s, that is

$$F^*(t|H_s^\pi)\mathbb{1}\{\tau_s \leq t\} = \mathbb{1}\{Y_s^\pi \in A \setminus \{Y_1^\pi, \ldots, Y_{s-1}^\pi\}\}\mathbb{1}\{\tau_s \leq t\} + F^*(t|H_{s+1}^\pi)\mathbb{1}\{\tau_s \leq t\}.$$

On the other hand if $\tau_s > t$, the behavior of OCL will be the same if played for $t-s$ steps from H_s^π or from H_{s+1}^π, that is

$$F^*(t-1|H_s^\pi)\mathbb{1}\{\tau_s > t\} = F^*(t|H_{s+1}^\pi)\mathbb{1}\{\tau_s > t\}.$$

Moreover note that

$$F^*(t-1|H_s^\pi) = F^*(t|H_s^\pi) - \mathbb{1}\{Y_t^*(H_s^\pi) \in A \setminus \{Y_1^\pi, \ldots, Y_{s-1}^\pi, Y_s^*(H_s^\pi), \ldots, Y_{t-1}^*(H_s^\pi)\}\}.$$

Thus we proved so far that

$$\begin{aligned}
&F^*(t|H_s^\pi) - F^*(t|H_{s+1}^\pi) \\
&= \mathbb{1}\{Y_s^\pi \in A \setminus \{Y_1^\pi, \ldots, Y_{s-1}^\pi\}\}\mathbb{1}\{\tau_s \leq t\} \\
&\quad + \mathbb{1}\{Y_t^*(H_s^\pi) \in A \setminus \{Y_1^\pi, \ldots, Y_{s-1}^\pi, Y_s^*(H_s^\pi), \ldots, Y_{t-1}^*(H_s^\pi)\}\}\mathbb{1}\{\tau_s > t\} \\
&\leq \mathbb{1}\{Y_s^\pi \in A \setminus \{Y_1^\pi, \ldots, Y_{s-1}^\pi\}\}\mathbb{1}\{\tau_s \leq t\} + \mathbb{1}\{Y_t^*(H_s^\pi) \in A \setminus \{Y_1^\pi, \ldots, Y_{s-1}^\pi\}\}\mathbb{1}\{\tau_s > t\}.
\end{aligned}$$

Now remark that Y_s^π is independent of τ_s conditionally to H_s^π. Thus one immediately obtains

$$\begin{aligned}
&\mathbb{E}[\mathbb{1}\{Y_s^\pi \in A \setminus \{Y_1^\pi,\dots,Y_{s-1}^\pi\}\}\mathbb{1}\{\tau_s \le t\}|H_s^\pi] \\
&= R_{I_s,n^\pi_{I_s,s-1}}\mathbb{E}[\mathbb{1}\{\tau_s \le t\}|H_s^\pi] \\
&\le R_{\bar{I}_s,n^\pi_{\bar{I}_s,s-1}}\mathbb{E}[\mathbb{1}\{\tau_s \le t\}|H_s^\pi].
\end{aligned}$$

Similarly $Y_t^*(H_s^\pi)$ is independent of $\mathbb{1}\{\tau_s > t\}$ conditionally to $(H_s^\pi, I_t^*(H_s^\pi))$ and thus

$$\begin{aligned}
&\mathbb{E}[\mathbb{1}\{Y_t^*(H_s^\pi) \in A \setminus \{Y_1^\pi,\dots,Y_{s-1}^\pi\}\}\mathbb{1}\{\tau_s > t\}|H_s^\pi, I_t^*(H_s^\pi)] \\
&= \mathbb{E}[R_{I_t^*(H_s^\pi),n^\pi_{I_t^*(H_s^\pi),s-1}}|H_s^\pi, I_t^*(H_s^\pi)]\mathbb{E}[\mathbb{1}\{\tau_s > t\}|H_s^\pi, I_t^*(H_s^\pi)] \\
&\le R_{\bar{I}_s,n^\pi_{\bar{I}_s,s-1}}\mathbb{E}[\mathbb{1}\{\tau_s > t\}|H_s^\pi, I_t^*(H_s^\pi)].
\end{aligned}$$

Putting everything together one obtains (12), which concludes the proof. ■

Lemma 9 *Let $a > 0$, $b \ge 0.4$, and $x \ge e$, such that $x \le a + b\log x$. Then one has*

$$x \le a + b\log\left(2a + 4b\log(4b)\right).$$

Proof If $a \ge b\log x$ then $x \le 2a$ and thus $x \le a + b\log(2a)$. On the other hand if $a < b\log x$ then $x \le 2b\log x$ which easily implies $x \le 4b\log(4b)$ (indeed for $x \ge e$, $x \mapsto \frac{x}{\log x}$ is increasing and furthermore for $b \ge 0.4$ one can check that $4b\log(4b) > 2b\log(4b\log(4b))$) and thus $x \le a + b\log(4b\log(4b))$. In any case one has $x \le a + b\log\left(2a + 4b\log(4b)\right)$. ■

Lemma 10 *For all $1 \le k \le n$,*

$$-\frac{1}{k} + \log\frac{n}{k} \le \sum_{j=k+1}^{n}\frac{1}{j} \le \log\frac{n}{k}\,.$$

Proof The standard sum/integral comparison yields

$$\log\frac{n+1}{k+1} \le \sum_{j=k+1}^{n}\frac{1}{j} \le \log\frac{n}{k}\,,$$

but

$$\log\frac{n+1}{k+1} = \log\frac{n}{k} + \log\left(1+\frac{1}{n+1}\right) - \log\left(1+\frac{1}{k+1}\right) \ge \log\frac{n}{k} + 0 - \frac{1}{k}\,.$$

■

Appendix B. The Open-loop Oracle Policy

In this final section, we provide an macroscopic analysis of the open-loop oracle policy in the case of uniform sampling, that is under Hypotheses (i), (ii) and (iii). An open-loop policy must choose, for each horizon t, the respective numbers of requests $(n_1^N,\dots,n_K^N)$ for each distribution (so that $n_1^N+\cdots+n_K^N=t^N$) in advance. It appears here that, in the limit, the *oracle open-loop* (OOL) policy, which makes use of the parameters $(Q_1^N,\dots,Q_K^N)$, is as good as the OCL policy.

Let here $\underline{R}^N_{i,n_i^N}=(Q_i^N-F_i^N(n_i^N))/N$ be the proportion of interesting items not yet found with expert i after n_i^N requests. Suppose that $t^N/N\to t$, and that $n_i^N/N\to\nu_i$ as N goes to infinity; it is easily shown that, almost surely,

$$\lim_{N\to\infty}\underline{R}^N_{i,n_i^N}=\lim_{N\to\infty}\mathbb{E}\left[\underline{R}^N_{i,n_i^N}\right]=\lim_{N\to\infty}\frac{Q_i^N\left(1-\frac{1}{N}\right)^{n_i^N}}{N}=q_i\exp(-\nu_i)\,.$$

Hence, the proportion of interesting items found with the allocation $(n_1^N,\dots,n_K^N)$ almost surely converges to $\sum_{i=1}^K q_i(1-\exp(-\nu_i))$. Defining

$$r(\nu)=\sum_{i=1}^K q_i\exp(-\nu_i)\,,$$

it follows that finding the best macroscopic allocation reduces to the following constrained convex minimization problem:

$$\min_{\nu\in\mathbb{R}^K} r(\nu)\quad\text{such that }\nu_1+\cdots+\nu_K=t\text{ and }\forall i,\,\nu_i\geq 0\,.$$

The solution $r^*(t)$, reached at $\nu=\nu^*(t)$, is easily derived by classical optimization techniques:

Proposition 11 *For every $i\in\{1,\dots,K\}$, let $\underline{q}_i=\exp\left(1/i\times\sum_{k=1}^i\log q_k\right)$ denotes the geometric mean of $q_1,\dots,q_i$.*

1. *There exists $I(t)\in\{1,\dots,K\}$ such that*

$$\begin{cases}\forall i\leq I(t), & \nu_i^*(t)=\frac{t}{I(t)}+\log\frac{q_i}{\underline{q}_{I(t)}}\\ \forall i>I(t), & \nu_i^*(t)=0\,.\end{cases}$$

   Hence,

$$r^*(t)=I(t)\underline{q}_{I(t)}\exp\left(-\frac{t}{I(t)}\right)+\sum_{i>I(t)}q_i\,.$$

2. *There exists $1=t_1\leq\cdots\leq t_K<+\infty$ such that*

$$\forall t\in[t_i,t_{i+1}[,\ I(t)=i\,.$$

   The $(t_k)_k$ are such that

$$q_i+(i-1)\underline{q}_{i-1}\exp\left(-\frac{t_i}{i-1}\right)=i\underline{q}_i\exp\left(-\frac{t_i}{i}\right)\,.$$

   For instance, $t_1=\log(q_1/q_2)$.

Proof: Introduce the Lagrangian:

$$L(\nu_1,\ldots,\nu_K,\lambda,\mu_1,\ldots,\mu_K) = \sum_{i=1}^{K} q_i \exp\left(-\frac{\nu_i}{N}\right) + \lambda\left(\sum_{i=1}^{K}\nu_i\right) - \sum_{i=1}^{K}\mu_i\nu_i\,.$$

We need to find the solution of:

$$\begin{gathered}\forall i\in\{1,\ldots,M\},\ \ -q_i\exp(-\nu_i)+\lambda-\mu_i=0,\\ \sum_{i=1}^{K}\nu_i = t,\\ \forall i\in\{1,\ldots,M\},\ \mu_i\nu_i=0 \text{ and } \mu_i\geq 0\,.\end{gathered}$$

We first obtain that

$$\nu_i = \log q_i - \log(\lambda-\mu_i)\,.$$

Denoting $A=\{i:\nu_i>0\}$, and using that $i\in A \implies \mu_i=0$, we get

$$t = \sum_{i\in A}\log(q_i) - |A|\log(\lambda)\,,$$

from which we get

$$-\log(\lambda) = \frac{t}{|A|} - \frac{1}{|A|}\sum_{i\in A}\log q_i,$$

and then for all $i\in A$:

$$\nu_i = \log q_i + \frac{t}{|A|} - \frac{1}{|A|}\sum_{i\in A}\log q_i\,.$$

Next, observe that $\nu_i=0 \iff q_i>\lambda$: in fact, if $\nu_i=0$ then the first equation gives $-q_i+\lambda-\mu_i=0$, and $0\leq\mu_i=\lambda-q_i$. Conversely, if $\nu_i>0$ then $\mu_i=0$ and $\nu_i=\log(q_i/\lambda)>0$ implies $q_i>\lambda$. Thus, there exists $I(t)$ such that $A=\{1,\ldots,I(t)\}$, and for all $i\leq I(t)$,

$$\nu_i = \log\frac{q_i}{\underline{q}_{I(t)}} + \frac{t}{I(t)}\,.$$

Moreover,

$$\begin{aligned} r^*(t) &= r\left(\nu_1,\ldots,\nu_{I(t)},0,\ldots,0\right)\\ &= \sum_{i\leq I(t)} q_i\exp\left[-\left(\log\frac{q_i}{\underline{q}_{I(t)}}+\frac{t}{I(t)}\right)\right] + \sum_{i>I(t)} q_i\\ &= I(t)\underline{q}_{I(t)}\exp\left(-\frac{t}{I(t)}\right) + \sum_{i>I(t)} q_i\,.\end{aligned}$$

The instants $(t_i)_{1\leq i\leq K}$ are such that

$$(i-1)\underline{q}_{i-1}\exp\left(-\frac{t_i}{i-1}\right) + \sum_{k>i-1} q_k = i\underline{q}_i\exp\left(-\frac{t_i}{i}\right) + \sum_{k>i} q_k\,,$$

which is equivalent to

$$q_i + (i-1)\underline{q}_{i-1}\exp\left(-\frac{t_i}{i-1}\right) = i\underline{q}_i\exp\left(-\frac{t_i}{i}\right) .$$

For $i = 2$, this gives

$$0 = q_2 + q_1\exp(-\nu_2) - 2\sqrt{q_1 q_2}\exp\left(-\frac{\nu_2}{2}\right) = \left(\sqrt{q_2} - \sqrt{q_1\exp(-\nu_2)}\right)^2 ,$$

which leads to $t_1 = \log(q_1/q_2)$.

Theorem 12 *In the macroscopic limit, the proportion of items found by the open-loop oracle policy uniformly converges to the function F defined in Equation* (7)*.*

The proportion of interesting items found by the OOL policy is

$$q - r^*(t) = \sum_{i\leq I(t)}\left[q_i - \underline{q}_{I(t)}\exp\left(-\frac{t}{I(t)}\right)\right] = \sum_{i=1}^{K}(q_i - \Lambda(t))_+ ,$$

where $\Lambda(t) = \underline{q}_{I(t)}\exp\left(-\frac{t}{I(t)}\right) \in [0, q_{I(t)}]$. To conclude, it remains only to remark that $\Lambda = T^{-1}$, where T is defined in Equation (6). In fact, if λ is such that $q_{i_0+1} < \lambda \leq q_{i_0}$, then $I(T(\lambda)) = i_0$ and

$$\Lambda(T(\lambda)) = \underline{q}_{i_0}\exp\left(-\frac{T(\lambda)}{i_0}\right) = \exp\left(\frac{1}{i_0}\sum_{i\leq i_0}\log q_i\right)\exp\left(-\frac{\sum_{i\leq i_0}\log(q_i/\lambda)}{i_0}\right) = \lambda .$$

If $\lambda < q_K$, the same holds with $i_0 = K$.

References

P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2):235–256, 2002.

S. Bubeck and N. Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122, 2012.

F. Fonteneau-Belmudes. *Identification of Dangerous Contingencies for Large Scale Power System Security Assessment*. PhD thesis, University of Liège, 2012.

F. Fonteneau-Belmudes, D. Ernst, C. Druet, P. Panciatici, and L. Wehenkel. Consequence driven decomposition of large-scale power system security analysis. In *Proceedings of the 2010 IREP Symposium - Bulk Power Systems Dynamics and Control - VIII*, Buzios, Rio de Janeiro, Brazil, August 2010.

W.A. Gale and G. Sampson. Good-turing frequency estimation without tears. *Journal of Quantitative Linguistics*, 2(3):217–237, 1995.

A. Garivier and O. Cappé. The KL-UCB algorithm for bounded stochastic bandits and beyond. In *Proceedings of the 24rd Annual International Conference on Learning Theory*, 2011.

I.J. Good. The population frequencies of species and the estimation of population parameters. *Biometrika*, 40:237–264, 1953. ISSN 0006-3444.

D. McAllester and L. Ortiz. Concentration inequalities for the missing mass and for histogram rule error. *J. Mach. Learn. Res.*, 4:895–911, December 2003. ISSN 1532-4435.

D.A. McAllester and R.E. Schapire. On the convergence rate of Good-Turing estimators. In *COLT*, pages 1–6, 2000.

C. McDiarmid. On the method of bounded differences. In *Surveys in combinatorics, 1989 (Norwich, 1989)*, volume 141 of *London Math. Soc. Lecture Note Ser.*, pages 148–188. Cambridge Univ. Press, Cambridge, 1989.

A. Orlitsky, N.P. Santhanam, and J. Zhang. Always good Turing: Asymptotically optimal probability estimation. In *FOCS '03: Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, pages 179+, Washington, DC, USA. IEEE Computer Society. ISBN 0-7695-2040-5.