

# Optimal discovery with probabilistic expert advice

Sébastien Bubeck  
Department of Operations and Financial Engineering  
Princeton University  
Princeton, NJ, 08544, USA  
sbubeck@princeton.edu

Damien Ernst  
Université of Liege  
Liegem B-4000, Belgium  
dernst@ulg.ac.be

Aurélien Garivier  
LTCI, CNRS & Telecom ParisTech  
Paris, France  
garivier@telecom-paristech.fr

October 26, 2011

## 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 the Good-Turing missing mass estimator. We show that this strategy uniformly attains the optimal discovery rate in a macroscopic limit sense, under some assumptions on the probabilistic experts. We also provide numerical experiments suggesting that this optimal behavior may still hold under weaker assumptions.

**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  $i^{\text{th}}$  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 be possible anymore since the computational resources required to simulate every contingency may exceed 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 [FB11, FBED<sup>+</sup>10] 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 shortcomings 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  $P_1, P_2, \dots, P_K$  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  $t$ , or equivalently to observe all the elements of  $A$  within as few time steps as possible. The index  $I_{t+1}$  may be chosen according to past observations: it is a (possibly randomized)

function of  $(I_1, X_{1,I_1}, \dots, I_t, X_{I_t, n_{I_t, t}})$ . We are mainly interested in the number of interesting items found by the strategy after  $t$  time steps:

$$\sum_{x \in A} \mathbb{1} \left\{ x \in \{X_{1,1}, \dots, X_{1, n_{1,t}}, \dots, X_{K,1}, \dots, X_{K, n_{K,t}}\} \right\}.$$

Note in particular that it of no interest to observe twice the same element of  $A$ .

While Algorithm Good-UCB, presented in Section 2, can be used in a more general setting (as illustrated in Section 6), for the mathematical analysis we restrict our attention to the case of probabilistic experts with the following properties:

- (i) non-intersecting supports:  $A \cap \text{supp}(P_i) \cap \text{supp}(P_j) = \emptyset$  for  $i \neq j$ ,
- (ii) finite supports with the same cardinality:  $|\text{supp}(P_i)| = N, \forall i \in \{1, \dots, K\}$ ,
- (iii) uniform distributions:  $P_i(x) = \frac{1}{N}, \forall x \in \text{supp}(P_i), \forall i \in \{1, \dots, K\}$ .

These assumptions are made in order to be able to compare the performance of the Good-UCB algorithm to an “oracle”, described below. Indeed, in that case, this oracle has a very simple behavior. In this setting it is convenient to reparametrize 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  $Q_i^N = |A^N \cap (\{i\} \times \{1, \dots, N\})|$  the number of interesting items accessible through requests to expert  $i$ . Further notation is given in Section 3.

### 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 proof of optimality in a particular setting. 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, [ACBF02]), and on a finite-time analysis of the Good-Turing estimator for the missing mass. In order to analyze and quantify the performance of this strategy, we compare it with the *oracle* (closed-loop) policy, a virtual algorithm that would be aware, at each time, of the probability of each item under each distribution, and would thus be able to sample optimally. This strategy is carefully described and analyzed in Section 3. The analysis is performed under the non-intersecting and uniform draws assumptions [(i), (ii), (iii)] described above, and in a macroscopic limit sense, that is when the size of the set  $\mathcal{X}$  grows to infinity while maintaining a constant proportion of interesting items. More precisely we prove the following theorem, where  $F^N(t)$  is the number of interesting items found by the oracle policy after  $t$  time steps.

**Theorem 1.** *Assume that, for all  $i \in \{1, \dots, K\}$ ,  $Q_i^N/N$  converges to  $q_i \in ]0, 1[$  as  $N$  goes to infinity. Then, almost surely, the sequence of mappings  $t \mapsto F^N([Nt])/N$  converges uniformly on  $\mathbb{R}_+$  to a limit denoted  $F$  as  $N$  goes to infinity.*

In Section 3 we also give an explicit expression for the limit  $F$ . Section 5 presents a study of the oracle *open-loop* policy which is defined as the optimal fixed allocation. In this problem it turns out that the oracle open-loop policy achieves the same performance as the oracle closed-loop policy, which in turns yields another formula for the macroscopic discovery rate  $F$ . In particular these formulas allow to see easily the difference in the macroscopic behavior

between optimal policies, and suboptimal policies such as uniform requests, see Remark 1 for more details.

The main result of the paper is given in Section 4. We show that Good-UCB is a macroscopically optimal policy, that is, the performances of Good-UCB tends to the performances of the oracle policy. More precisely let  $\tilde{F}^N(t)$  be the number of interesting items found by Good-UCB after  $t$  time steps.

**Theorem 2.** *Assume that, for all  $i \in \{1, \dots, K\}$ ,  $Q_i^N/N$  converges to  $q_i \in ]0, 1[$  as  $N$  goes to infinity. Then, almost surely, the sequence of mappings  $t \mapsto \tilde{F}^N([Nt])/N$  converges uniformly on  $\mathbb{R}_+$  to the limiting proportion  $F$  found during the same time by the oracle policy.*

Section 6 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 anymore. Finally, Section 7 concludes.

## 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 [Agr95, ACBF02] 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 with the Good-Turing estimator and a brief study of its concentration properties. Then we describe precisely the Good-UCB strategy.

### 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  $\Omega$  be a discrete set, and let  $A$  be a subset of interesting elements of  $\Omega$ . Assume that  $X_1, \dots, X_n$  are elements of  $\Omega$  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  $p_{\max} = \max\{P(x) : x \in \Omega\}$ , let  $R_n = \sum_{x \in A} Z_n(x)P(x)$  denote the missing mass of the interesting items, and let  $U_n = \sum_{x \in A} U_n(x)$  the number of elements of  $A$  that have been seen exactly once (in linguistics, they are often called *appaxes*). The idea of the Good-Turing estimator ([Goo53], see also [MS00, OSZ03] 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  $X_1, \dots, X_n$ , by the “fraction of appaxes”  $\hat{R}_n = U_n/n$ . This estimator is well-known in linguistics, for instance in order to estimate the number of words in some language,

see [GS95]. For our particular needs, we derive (using similar techniques as in [MS00]) the following upper-bound on the estimation error:

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

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

**Proof:** The random variable  $W_n = R_n - \hat{R}_n$  is a function of the independent observations  $X_1, \dots, X_n$  such that, denoting  $W_n = f(X_1, \dots, X_n)$ , modifying just one observation has limited impact:  $\forall l \in \{1, \dots, n\}, \forall (x_1, \dots, x_n, x'_l) \in \Omega^{n+1}$ ,

$$|f(x_1, \dots, x_n) - f(x_1, \dots, x_{l-1}, x'_l, x_{l+1}, \dots, x_n)| \leq \frac{2}{n} + p_{\max}$$

By applying McDiarmid's inequality [McD89], one gets that, with probability at least  $1 - \delta$ ,

$$|W_n - \mathbb{E}[W_n]| \leq \sqrt{\frac{(2/n + p_{\max})^2 n \log(2/\delta)}{2}}.$$

Moreover,

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

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 estimated procedure it relies on. For every arm  $i \in \{1, \dots, K\}$  and for every  $t \in \mathbb{N}$ , denote

$$O_{i,t}(x) = \sum_{s=1}^{n_{i,t}} \mathbb{1}\{X_{i,s} = x\}, \quad O_t(x) = \sum_{i=1}^K O_{i,t}(x), \quad U_{i,t}(x) = \mathbb{1}\{O_{i,t}(x) = O_t(x) = 1\}, \\ U_{i,t} = \sum_{x \in A} U_{i,t}(x).$$

For each arm  $i \in \{1, \dots, K\}$ , the index at time  $t$  is composed of the estimate

$$\hat{R}_{i,t-1} = \frac{U_{i,t-1}}{n_{i,t-1}}$$

of the missing mass

$$\sum_{x \in A \setminus \{X_{I_1, n_{I_1, 1}}, \dots, X_{I_{t-1}, n_{I_{t-1}, t-1}}\}} P_i^N(x)$$

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.

Note that the Good-UCB algorithm is designed for more general probabilistic experts than those satisfying assumptions [(i), (ii), (iii)]. In particular since we do not make the non-intersecting supports assumption (i), the missing mass of a given expert  $i$  depends explicitly on

---

**Algorithm 1** Good-UCB

---

- 1: For  $1 \leq t \leq K$  choose  $I_t = t$ .
  - 2: **for**  $t \geq K + 1$  **do**
  - 3:   Choose  $I_t = \arg \max_{1 \leq i \leq K} \left\{ \hat{R}_{i,t-1} + c \sqrt{\frac{\log(t)}{n_{i,t-1}}} \right\}$
  - 4:   Observe  $X_t$  distributed as  $P_{I_t}$  and update  $\hat{R}_{1,t}, \dots, \hat{R}_{K,t}$  accordingly
  - 5: **end for**
- 

the outcomes of *all* requests (and not only requests to expert  $i$ ). Note also that the bounds of Proposition 1 hold for all discrete distributions. The experiments of Section 6 validate these observations, and show that Good-UCB behaves very well even when assumptions [(i), (ii), (iii)] are not met. However, for the theoretical analysis of our algorithm, we focus on large values of  $N$  under the non-intersecting and uniform draws assumptions [(i), (ii), (iii)]: indeed in that case the performance of the oracle strategy is simple and deterministic, so that the optimality of the Good-UCB algorithm can be analyzed. More precisely, we will show that, in the macroscopic limit, the number of items found at each time by Good-UCB converges to the number of items found by the closed-loop oracle strategy that knows the number of interesting items to find with each expert, at every time, and that may use this information to make its choice. In order to prove this, we first analyze the performance of such an oracle strategy.

### 3 The closed-loop oracle strategy

From now on we restrict our attention to the setting described in Section 1.2 with sets  $\mathcal{X}^N, A^N$ , and experts distributions  $P_1^N, \dots, P_K^N$ . Denote by  $B_i^N$  the set of interesting items supported by  $P_i^N$ :  $B_i^N = \{x \in \{1, \dots, N\} : (i, x) \in A^N\}$ . Let  $Q_i^N = |B_i^N|$ ; in particular, note that  $Q^N = Q_1^N + \dots + Q_K^N$ . Without loss of generality, we will assume in the analysis that  $Q_1^N \geq Q_2^N \geq \dots \geq Q_K^N$ . 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. We denote by  $(D_{i,k}^N)_{1 \leq k \leq Q_i^N}$  the increasing sequence of the indices corresponding to draws for which new interesting items are discovered with expert  $i$ :

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

Reciprocally, we denote  $F_{i,n}^N = \max\{k \in \mathbb{N} : D_{i,k}^N \leq n\}$  the number of items found in the first  $n$  draws. We also define  $S_{i,0}^N = 0$  and for  $k \geq 1$ ,  $S_{i,k}^N = D_{i,k}^N - D_{i,k-1}^N$ . The random variables  $S_{i,k}^N$  ( $1 \leq i \leq K, k \geq 1$ ) are independent with geometric distribution  $\mathcal{G}((1 + Q_i^N - k)/N)$ . In particular, for all  $k \geq 1$ ,

$$\mathbb{E} [D_{i,k}^N] = \frac{N}{Q_i^N} + \dots + \frac{N}{1 + Q_i^N - k}. \quad (1)$$

#### 3.1 Description of the closed-loop oracle policy

When the values of  $Q_1^N, \dots, Q_K^N$  are known, so that the number of interesting items to find with each expert is known at every step, an horizon-free optimal closed-loop strategy (denoted in the following as the “oracle closed-loop strategy” or as OCL) consists in making a request, at each time step, to one of the experts with highest number of still undiscovered interesting items. Hence, an OCL strategy can:

- first request expert 1 for  $D_{1, Q_1^N - Q_2^N}^N$  steps;

- then, alternatively request
  - expert 1 for  $S_{1,1+Q_1^N-Q_2^N}^N$  steps;
  - expert 2 for  $S_{2,1}^N$  steps;
  - expert 1 for  $S_{1,2+Q_1^N-Q_2^N}^N$  steps;
  - expert 2 for  $S_{2,2}^N$  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,  $\dots$ ,  $K$  in the alternance.

For every  $l \in \{0, \dots, Q_1^N\}$ , we shall be particularly interested in the waiting time  $T^N(l)$  until, under the OCL strategy described above, all experts have at most  $l$  undiscovered interesting items. Obviously,

$$T^N(l) = \sum_{i:Q_i^N>l} D_{i,Q_i^N-l}^N,$$

so that, by Equation (1),

$$\mathbb{E} [T^N(l)] = \sum_{i:Q_i^N>l} \frac{N}{Q_i^N} + \dots + \frac{N}{l+1}. \quad (2)$$

At that time, the number of items discovered so far is  $G^N(l) = \sum_{i=1}^K (Q_i^N - l)_+$ . For every  $f \in \{0, \dots, Q^N\}$ , we define  $L^N(f)$  to be the maximal number of undiscovered items remaining on an expert, once that  $f$  interesting items have been discovered altogether. In particular, note that  $L^N(G^N(l)) = l$  for all  $l$ , and that  $G^N(L^N(f)) \in \{f - K + 1, \dots, f\}$  for all  $f$ . Besides, the time  $\tau^N(f)$  required by the OCL strategy to collect  $f$  interesting items satisfies

$$T^N(L^N(f)) \leq \tau^N(f) < T^N(L^N(f) - 1). \quad (3)$$

The performance of the OCL strategy is maybe more explicitly expressed by the pseudo-inverse of the mapping of  $\tau^N$ : for every integer  $t$  let  $F^N(t)$  denote the total number of items found up to time  $t$ .  $\tau^N$  and  $F^N$  are related as follows:

$$\begin{aligned} \forall t \geq 0, F^N(t) &= \max\{f \in \{1, \dots, Q^N\} : \tau^N(f) \leq t\}, \\ \forall f \in \{0, \dots, Q^N\}, \tau^N(f) &= \min\{t \geq 0 : F^N(t) = f\}. \end{aligned} \quad (4)$$

### 3.2 Macroscopic limit

We consider a macroscopic limit where  $N$  goes to infinity together with the  $Q_i^N$  in such a way that  $Q_i^N/N \rightarrow q_i \in ]0, 1[$ . Let  $q = q_1 + \dots + q_K$ . We will show that if  $f^N$  is a sequence of integers such that  $f^N/N \rightarrow \phi \in ]0, q[$ , then the normalized waiting time  $\tau^N(f^N)/N$  converges as  $N$  goes to infinity to a deterministic limit  $\tau(\phi)$  that we will compute as a function of  $\phi$  and  $q_1, \dots, q_K$ . We start with some notation. Let  $G : [0, q_1] \rightarrow [0, q]$  be the strictly decreasing mapping defined by

$$G(\lambda) = \sum_{i=1}^K (q_i - \lambda)_+,$$

and let  $L : [0, q] \rightarrow [0, q_1]$  be the inverse mapping. The following lemmas are proved in the appendix. The second lemma is the key step of the macroscopic analysis.



**Lemma 1.** *The mappings  $G$  and  $L$  are the limits of the sequences of mappings  $(G^N)_N$  and  $(L^N)_N$ , respectively, in the following sense:*

- (i) *If  $l^N \in [0, Q_1^N]$  defines a sequence of integers, then  $l^N/N \rightarrow \lambda \in [0, q_1]$  if and only if  $G^N(l^N)/N \rightarrow G(\lambda)$  as  $N$  goes to infinity.*
- (ii) *If  $f^N \in [0, Q^N]$  defines a sequence of integers, then  $f^N/N \rightarrow \phi \in [0, q]$  if and only if  $L^N(f^N)/N \rightarrow L(\phi)$  as  $N$  goes to infinity.*

**Lemma 2.** *For every  $\lambda \in ]0, q_1]$ , let*

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

*For every sequence  $(l^N)_N$  such that  $l^N/N$  converges to  $\lambda$  as  $N$  goes to infinity,  $T^N(l^N)/N$  converges almost surely to  $T(\lambda)$  as  $N$  goes to infinity.*

**Theorem 3.** *For every sequence  $(f^N)_N$  such that  $f^N \in \{0, \dots, Q^N\}$  and  $f^N/N$  converges to  $\phi$  as  $N$  goes to infinity,  $\tau^N(f^N)/N$  converges a.s. to  $\tau(\phi) = T(L(\phi))$  as  $N$  goes to infinity.*

**Proof:** By Equation (3),

$$\frac{T^N(L^N(f^N))}{N} \leq \frac{\tau^N(f^N)}{N} < \frac{T^N(L^N(f^N) - 1)}{N}.$$

By Lemma 1,  $\lim_{N \rightarrow \infty} L^N(f^N)/N = \lim_{N \rightarrow \infty} (L^N(f^N) - 1)/N = L(\phi)$  and thus, by Lemma 2,

$$\lim_{N \rightarrow \infty} \frac{T^N(L^N(f^N))}{N} = \lim_{N \rightarrow \infty} \frac{T^N(L^N(f^N) - 1)}{N} = T(L(\phi)).$$

**Corollary 1.** *For every sequence sequence  $t^N$  of integers such that  $t^N/N \rightarrow t$  as  $N$  goes to infinity,  $F^N(t^N)$  converges to  $F(t)$ , where the function  $F: \mathbb{R} \rightarrow [0, q[$  is  $\tau^{-1} = G \circ T^{-1}$ , i.e.*

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

The proof, very similar to the previous proofs, is omitted. Another expression for  $F$  is obtained in Section 5 in the analysis of the *open-loop* oracle policy. This allows us to finish the proof of Theorem 1:

**Proof of Theorem 1 :** The fact that, almost surely, the sequence of increasing processes  $t \mapsto F^N([Nt])/N$  uniformly converges to  $F$  on every compact of  $\mathbb{R}_+$  is a consequence of Corollary 1 and Dini's Theorem. This is sufficient, since  $F$  is upper-bounded by  $q$ .

## 4 Macroscopic optimality of the Good-UCB algorithm

After  $n$  requests to expert  $i \in \{1, \dots, K\}$ , denote by  $\underline{R}_{i,n}^N = (Q_i^N - F_i^N(n))/N$  the proportion of interesting items not yet found with that expert. To estimate this number, we use the Good-Turing estimator

$$\hat{\underline{R}}_{i,n}^N = \frac{\sum_{x \in B_i^N} \mathbb{1}\{\sum_{m=1}^n \mathbb{1}\{X_{i,m} = x\} = 1\}}{n}.$$



In particular note that under assumption (i), the estimator  $\hat{R}_{i,t}$  defined in Section 2.2 satisfies  $\hat{R}_{i,t} = \hat{R}_{i,n_{i,t}}$ . To simplify the proof, we use here a slightly different confidence bonus than the one proposed on line 3 of the algorithm. We define here the following upper confidence bound:

$$u_{i,n}^N = \hat{R}_{i,n}^N + \sqrt{\frac{(2/n + 1/N)^2 n \log(2KN^4)}{2}}.$$

According to Proposition 1, the event  $C^N$  such that:

$$\forall i \in \{1, \dots, K\}, \forall n \in \{1, N^2\}, u_{i,n}^N - 2\sqrt{\frac{(2/n + 1/N)^2 n \log(2KN^4)}{2}} - \frac{1}{n} \leq \underline{R}_{i,n}^N \leq u_{i,n}^N, \quad (5)$$

has probability at least  $1 - N^{-2}$ .

The Good-Turing optimistic strategy consists in making a request, at each step, to the expert maximizing the upper-confidence bound  $u_{i,n_{i,t}}^N$ , where  $n_{i,t}$  denotes the (random) number of requests to expert  $i$  before time  $t$ . Denote by  $\tilde{\tau}^N(f)$  the time required by this algorithm to collect  $f$  interesting items, and by  $\tilde{F}^N(t)$  the number of interesting items found by Good-UCB in the first  $t$  rounds.

**Theorem 4.** *In the macroscopic limit,  $\tilde{\tau}^N(f^N)/N$  converges a.s. to the same limit  $\tau(\phi)$  as for the oracle closed-loop policy when  $f^N/N$  tends to  $\phi$ . Moreover,  $\tilde{F}^N(t^N)/N$  converges a.s. to the limiting proportion  $F(t)$  found during the same time by the OCL policy when  $t^N/N$  converges to  $t$ .*

As for Theorem 1, this results directly leads to Theorem 2. To prove Theorem 4, we proceed as for Theorem 3: we first consider for every  $\epsilon > 0$  and for every  $l \in \{\epsilon N, \dots, Q_1^N\}$  the number  $\tilde{T}^N(l)$  of steps until, using the Good-UCB algorithm, at most  $l$  interesting items remain undiscovered on all experts. Obviously, on  $C^N$ , the number  $\tilde{T}^N(l)$  is upper-bounded by the first time

$$U^N(l) = \inf \left\{ t > 0 : \forall i \in \{1, \dots, K\}, u_{i,n_{i,t}}^N \leq \frac{l}{N} \right\}$$

when all the upper-bounds of the missing masses are below  $l/N$ . Write

$$U^N(l) = U_1^N(l) + \dots + U_K^N(l),$$

where  $U_i^N(l) = n_{i,U^N(l)}$  denotes the number of requests to expert  $i \in \{1, \dots, K\}$  up to time  $U^N(l)$ . Four cases can be distinguished:

- either the draw does not belong to  $C^N$ , which has probability  $O(N^{-2})$ ;
- or  $U^N(l) \geq N^2$ , which is easily shown to have probability  $O(N^{-2})$ ; in fact, after  $6N \log(N)$  requests to an expert  $i$  the probability that there exists any item  $(i, x) \in \Omega^N$  which has not been drawn at least twice is  $O(N^{-2})$ , and otherwise  $u_{i,N^{3/2}}^N = O\left(\sqrt{\log(N)/N}\right) < \epsilon$  for  $N$  large enough;
- or  $U_i^N(l) \leq N^{2/3}$ ;
- or, as  $N^{2/3} \leq U_i^N(l) \leq U^N(l) \leq N^2$ , and according to Equation 5,

$$\begin{aligned} \underline{R}_{i,U_i^N(l)}^N &\geq u_{i,U_i^N(l)}^N - 2\sqrt{\frac{(2/U_i^N(l) + 1/N)^2 U_i^N(l) \log(2KN^4)}{2}} - \frac{1}{U_i^N(l)} \\ &\geq u_{i,U_i^N(l)}^N - CN^{-2/3} \log(KN), \end{aligned}$$

for some positive absolute constant  $C$ . Now, remark that

- $u_{i,U_i^N(l)-1}^N > l$ , because the  $(U_i^N(l) - 1)$ -th request to expert  $i$  took place before time  $U^N(l)$ , and at that time expert  $i$  had the highest index;
- $u_{i,U_i^N(l)}^N \geq u_{i,U_i^N(l)-1}^N - C'N^{-2/3}$  for some absolute constant  $C'$ , as  $n \geq N^{2/3}$ ,  $u_{i,n}^N - u_{i,n-1}^N \geq -1/n$  and  $\sqrt{(2/n + 1/N)^2 n \log(2KN^4)/2} - \sqrt{(2/(n-1) + 1/N)^2 (n-1) \log(2KN^4)/2} \geq -1/n$  because

$$\frac{\sqrt{(2/n + 1/N)^2 n \log(2KN^4)/2}}{\sqrt{(2/(n-1) + 1/N)^2 (n-1) \log(2KN^4)/2}} \geq \sqrt{\frac{n}{n+1}} \geq 1 - \frac{1}{n},$$

and  $\sqrt{(2/(n-1) + 1/N)^2 (n-1) \log(2KN^4)/2} \leq 1$  for  $n \geq KN^{3/5}$ .

Hence  $\underline{R}_{i,U_i^N(l)}^N \geq l - \Delta^N/N$ , with  $\Delta^N \leq C''N^{1/3} \log(KN)$  and  $C'' = C + C'$ . As  $\underline{R}_k^N(i) = (Q_i^N - F_{i,k}^N)/N$  for every positive integer  $k$ , this implies that  $U_i^N(l) \leq D_{i,Q_i^N-l-\Delta^N}^N$ .

Altogether, we obtain that on  $C^N$ :

$$U^N(l) \leq KN^{2/3} + \sum_{i:Q_i^N > l - \Delta^N} D_{i,Q_i^N-l-\Delta^N}^N = KN^{2/3} + T^N(l - \Delta^N).$$

Thus, if  $l^N$  is a sequence of integers of  $[\epsilon N, Q_1^N]$  such that  $l^N/N \rightarrow \lambda \in [\epsilon, q_1]$  as  $N$  goes to infinity, then

$$\limsup_{N \rightarrow \infty} \frac{U^N(l^N)}{N} \leq \lim_{N \rightarrow \infty} \frac{T^N(l^N - \Delta^N)}{N} = T(\lambda)$$

according to Lemma 2 almost surely, using the Borel-Cantelli lemma and the fact that  $\sum_N P(C^N \cup \{U^N(l) \geq N^2\}) < \infty$ . This is sufficient to show that  $\tilde{T}(l^N)/N$  converges a.s. to  $T(\lambda)$  as  $N$  goes to infinity. As for Corollary 1, this implies that  $\tilde{F}^N(t^N)/N$  converges a.s. to  $F(t)$  when  $t^N/N$  converges to  $t$ .

## 5 The open-loop oracle policy

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  $n_1^N + \dots + n_K^N = t^N$ ) 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.

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

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

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:

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

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

**Proposition 2.** 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  $1 = t_1 \leq \dots \leq t_K < +\infty$  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:

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

We need to find the solution of:

$$\forall i \in \{1, \dots, M\}, \quad -q_i \exp(-\nu_i) + \lambda - \mu_i = 0$$

$$\sum_{i=1}^K \nu_i = t$$

$$\forall i \in \{1, \dots, M\}, \quad \mu_i \nu_i = 0 \text{ and } \mu_i \geq 0$$

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  $A = \{1, \dots, I(t)\}$ , and for all  $i \leq I(t)$ ,

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

Moreover,

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

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 5.** *The proportion of items found by the open-loop oracle policy uniformly converges to  $F$  in the macroscopic limit.*

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  $\Lambda = T^{-1}$ : in fact, if  $\lambda$  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  $i_0 = K$ .

**Remark 1.** *Proposition 2 permits to compare easily the macroscopic performance of the Good-UCB algorithm with balanced sampling: when all distributions are sampled equally often, the proportion of unseen interesting items at time  $t$  is  $\sum_{i=1}^K q_i \exp(-t/K) = K\bar{q}_K \exp(-t/K)$ , with  $\bar{q}_K = (\sum_{i=1}^K q_i)/K$ . On the other side, for the Good-UCB algorithm, this proportion is  $I(t)\underline{q}_{I(t)} \exp(-t/I(t))$ . The ratio is a decreasing function of time, but even when  $t > t_K$ , it takes the nice expression  $\bar{q}_K/\underline{q}_K \geq 1$ , the ratio of an arithmetic mean with a geometric mean, which is (as expected) as high as the  $(q_i)_i$  are unbalanced.*

## 6 Simulations

We provide a few simulations illustrating the behaviour of the Good-UCB algorithm in practice. In order to illustrate the convergence properties shown in Sections 3 and 4, 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  $q_7 = 0.8\%$ . Figure 1 displays the number of items found as a function of time by the Good-UCB (solid), the oracle (dashed) and a balanced sampling scheme simply alternating 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). The convergences proved in Corollary 1 and Theorem 4 are obvious. 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 oracle,

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.

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 policy, Good-UCB and uniform sampling. The results are displayed in Figure 2. Even if the difference remains significant between Good-UCB and the oracle, 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, one can 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.

## 7 Conclusions

This paper introduced an original problem, optimal discovery with probabilistic expert advice. We proposed an algorithm to solve this problem, and showed both analytically and through simulations its efficiency.

This work can be extended along several directions. First, it would be interesting to analyze the behaviour of Good-UCB under less restrictive assumptions on the experts. Note that

Figure 1: Number of items found by Good-UCB (solid), the oracle (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.

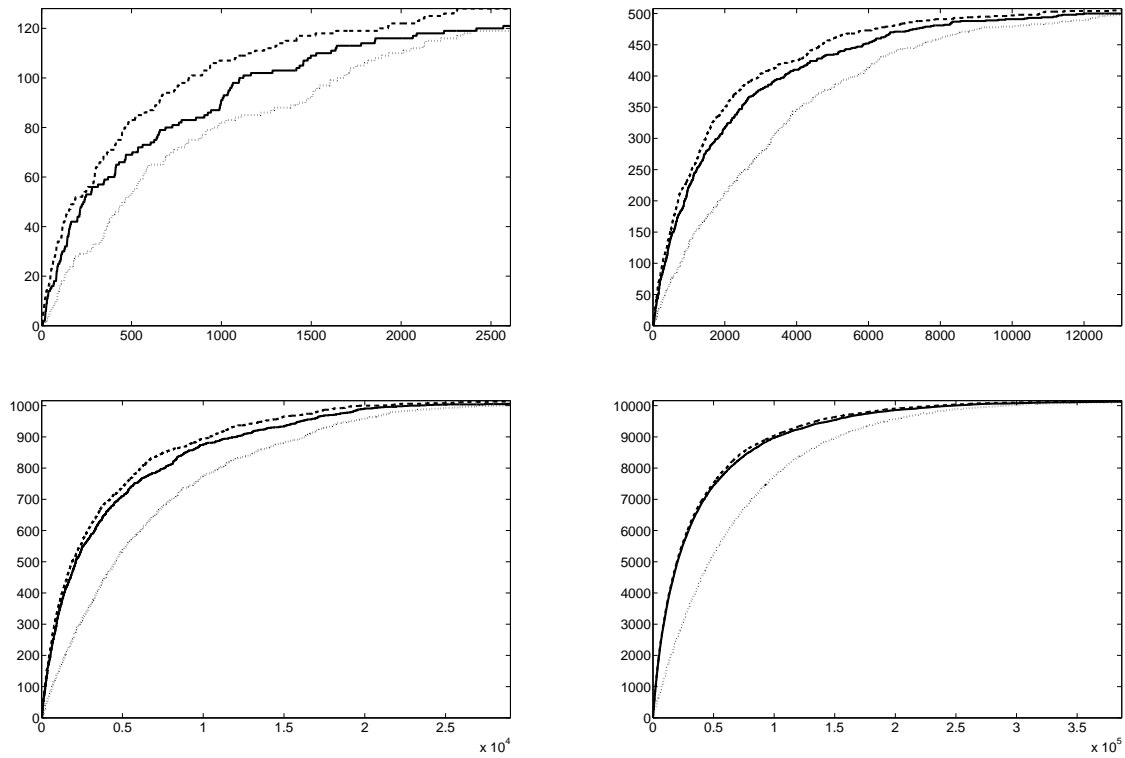
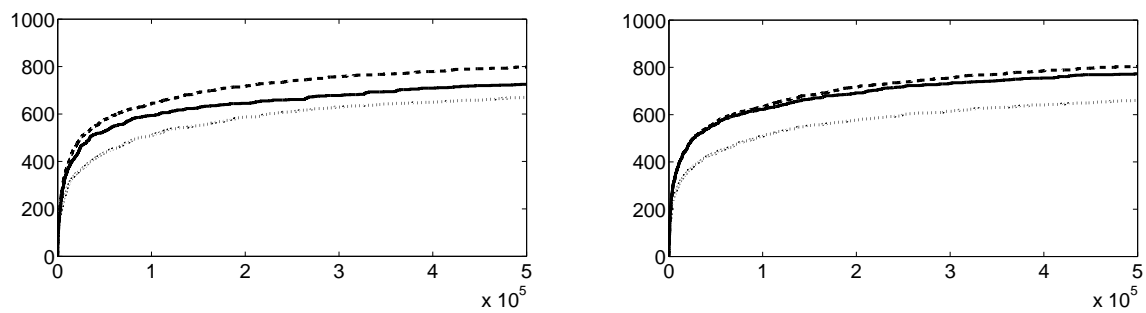


Figure 2: Number of prime numbers found by Good-UCB (solid), the oracle (dashed), and uniform sampling (dotted) as a function of time, using geometric experts with means 100, 300, 500, 700 and 900, for  $c = 0.1$  (left) and  $c = 0.02$  (right).



assumptions (ii) and (iii) are used only to (considerably) simplify the analysis of the oracle policy, and hence to prove the optimality of Good-UCB. Removing assumption (ii) seems fairly straightforward up to the addition of another level of notations. Because of the intricate behavior of oracle policies in that case, it is not clear how the analysis could be carried out if assumption (iii) were to be removed, though it seems reasonable to assume that Good-UCB will still be macroscopically optimal. Good-UCB is designed to work even when assumption (i) is not satisfied, but the analysis is complicated by the explicit dependency between the missing mass of the different experts.

One may also wonder whether it would be possible to obtain optimal *rates of convergence* (in the macroscopic limit sense) for this problem, and whether Good-UCB is optimal in that sense too. Finally, another macroscopic limit deserves to be investigated, where the number of interesting items for each arm remains constant, while  $N$  and  $n$  go to infinity. Note that in such a case, a Poisson regime appears. The analysis of Good-UCB might be possible by using a better concentration bound for the Good-Turing estimator such as the Boucheron-Massart-Lugosi inequality [BLM09]. This could also contribute to explain why, in the second experiment presented in Section 6, the parameter  $c$  should be chosen decreasing with time.

## 8 Appendix

**Lemma 3.** For all  $1 \leq k \leq n$ ,

$$-\frac{1}{k} + \log \frac{n}{k} \leq \sum_{j=k+1}^n \frac{1}{j} \leq \log \frac{n}{k} \quad (6)$$

**Proof:** The standard sum/integral comparison yields

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

but

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

**Proof of Lemma 1:** For (i), it suffices to notice that

$$\frac{1}{N} G^N(l^N) = \sum_{i=1}^K \left( \frac{Q_i^N}{N} - \frac{l^N}{N} \right)_+ \rightarrow \sum_{i=1}^K (q_i - \lambda)_+$$

if  $l^N/N \rightarrow \lambda$ . Moreover, the same argument shows that

$$\limsup \frac{1}{N} G^N(l^N) = G \left( \limsup \frac{l^N}{N} \right) \quad \text{and} \quad \liminf \frac{1}{N} G^N(l^N) = G \left( \liminf \frac{l^N}{N} \right),$$

Hence, if  $G^N(l^N)/N$  converges, the limit belongs to  $[0, q]$  and thus can be written  $G(\lambda)$  for some  $\lambda \in [0, q_1]$ . Thus,  $G \left( \liminf \frac{l^N}{N} \right) = G \left( \limsup \frac{l^N}{N} \right) = G(\lambda)$ , which implies that  $l^N/N \rightarrow \lambda$  as  $G$  a continuous bijection.

Concerning (ii): if  $f^N/N \rightarrow \phi$ , then the fact that  $|G^N(L^N(f^N)) - f^N| < K$  implies that  $G^N(L^N(f^N))/N \rightarrow \phi$  and thus, by (i), that  $L^N(f^N)/N$  converges to a value  $\lambda$  such that  $G(\lambda) = \phi$ , i.e.  $\lambda = L(\phi)$ . The reciprocal (which is not used in the sequel) is left to the reader.



**Proof of Lemma 2:** Since

$$T^N(l^N) = \sum_{i:Q_i^N > l^N} D_{i,Q_i^N-l^N}^N,$$

it suffices to show that for every expert  $i \in \{1, \dots, K\}$ ,  $D_{i,Q_i^N-l^N}^N/N$  converges almost surely to  $\log(q_i/\lambda)$  as  $N$  goes to infinity. Write

$$W_{i,l^N}^N = \frac{1}{N} \left( D_{i,Q_i^N-l^N}^N - \mathbb{E} \left[ D_{i,Q_i^N-l^N}^N \right] \right) = \frac{1}{N} \sum_{k=1}^{Q_i^N-l^N-1} (S_{i,k}^N - \mathbb{E} [S_{i,k}^N]). \quad (7)$$

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

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

for some positive constant  $c(d)$  depending only on  $d$ , and for  $N$  large enough. Hence, taking (7) to the fourth power and developing yields

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

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

$$\frac{1}{N} \mathbb{E} \left[ D_{i,Q_i^N-l^N}^N \right] = \frac{1}{Q_1^N} + \dots + \frac{1}{l^N + 1} = \log \frac{Q_i^N}{l^N} - \epsilon^N,$$

with  $0 \leq \epsilon^N \leq 1/l^N$  according to Lemma 3, and thus

$$\frac{1}{N} \mathbb{E} \left[ D_{i,Q_i^N-l^N}^N \right] \rightarrow \lim_{N \rightarrow \infty} \log \left( \frac{Q_i^N/N}{l^N/N} \right) = \log(q_i/\lambda),$$

which concludes the proof.

## References

- [ACBF02] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2):235–256, 2002.
- [Agr95] R. Agrawal. Sample mean based index policies with  $O(\log n)$  regret for the multiarmed bandit problem. *Advances in Applied Probability*, 27(4):1054–1078, 1995.
- [BLM09] S. Boucheron, G. Lugosi, and P. Massart. On concentration of self-bounding functions. *Electron. J. Probab.*, 14:no. 64, 1884–1899, 2009.
- [FB11] F. Fonteneau-Belmudes. *Identification of dangerous contingencies for large scale power system security assessment*. PhD thesis, University of Liège, 2011.
- [FBED<sup>+</sup>10] 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.

- [Goo53] I.J. Good. The population frequencies of species and the estimation of population parameters. *Biometrika*, 40:237–264, 1953.
- [GS95] W.A. Gale and G. Sampson. Good-turing frequency estimation without tears. *Journal of Quantitative Linguistics*, 2(3):217–237, 1995.
- [McD89] 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.
- [MS00] D.A. McAllester and R.E. Schapire. On the convergence rate of good-turing estimators. In *COLT*, pages 1–6, 2000.
- [OSZ03] 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, 2003. IEEE Computer Society.