A KL-LUCB algorithm for Large-Scale Crowdsourcing

Authored by:

Robert Nowak
Ervin Tanczos
Bob Mankoff

Abstract

This paper focuses on best-arm identification in multi-armed bandits with bounded rewards. We develop an algorithm that is a fusion of lil-UCB and KL-LUCB, offering the best qualities of the two algorithms in one method. This is achieved by proving a novel anytime confidence bound for the mean of bounded distributions, which is the analogue of the LIL-type bounds recently developed for sub-Gaussian distributions. We corroborate our theoretical results with numerical experiments based on the New Yorker Cartoon Caption Contest.

1 Paper Body

This paper focuses on best-arm identification in multi-armed bandits with bounded rewards. We develop an algorithm that is a fusion of lil-UCB and KL-LUCB, offering the best qualities of the two algorithms in one method. This is achieved by proving a novel anytime confidence bound for the mean of bounded distributions, which is the analogue of the LIL-type bounds recently developed for sub-Gaussian distributions. We corroborate our theoretical results with numerical experiments based on the New Yorker Cartoon Caption Contest.

1

Multi-Armed Bandits for Large-Scale Crowdsourcing

This paper develops a new multi-armed bandit (MAB) for large-scale crowdsourcing, in the style of the KL-UCB [4, 9, 3]. Our work is strongly motivated by crowdsourcing contests, like the New Yorker Cartoon Caption contest [10]3 . The new approach targets the ?best-arm identification problem? [1] in the fixed confidence setting and addresses two key limitations of existing theory and algorithms: (i) State of the art algorithms for best arm identification are based on sub-Gaussian confidence bounds [5] and fail to exploit the fact that rewards are usually bounded in crowdsourcing applications. (ii) Existing KL-UCB algorithms for best-arm identification do exploit bounded rewards [8] , but

have suboptimal performance guarantees in the fixed confidence setting, both in terms of dependence on problem-dependent hardness parameters (Chernoff information) and on the number of arms, which can be large in crowdsourcing applications. The new algorithm we propose and analyze is called lil-KLUCB, since it is inspired by the lil-UCB algorithm [5] and the KL-LUCB algorithm [8]. The lil-UCB algorithm is based on sub-Gaussian bounds and has a sample complexity for best-arm identification that scales as X ?1 ??2 log ??2 i log(? i), i?2

where ? ? (0, 1) is the desired confidence and ?i = ?1 ? ?i is the gap between the means of the best arm (denoted as arm 1) and arm i. If the rewards are in [0, 1], then the KL-LUCB algorithm has ? 3

This work was partially supported by the NSF grant IIS-1447449 and the AFSOR grant FA9550-13-1-0138. For more details on the New Yorker Cartoon Caption Contest, see the Supplementary Materials.

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

a sample complexity scaling essentially like4 X (Di?)?1 log(n? ?1 (Di?)?1) , i?2

where n is the number of arms and Di? := D? (?1 , ?i) is the Chernoff-information between a Ber(?1) and a Ber(?i) random variable5 . Ignoring the logarithmic factor, this bound is optimal for the case of Bernoulli rewards [7, 11]. Comparing these two bounds, we observe that KL-LUCB may offer benefits since Di? = D? (?1 , ?i) ? (?1 ? ?i)2 /2 = ?2i /2, but lil-UCB has better logarithmic dependence on the ?2i and no explicit dependence on the number of arms n. Our new algorithm lil-KLUCB offers the best of both worlds, providing a sample complexity that scales essentially like X (Di?)?1 log(? ?1 log(Di?)?1) . i?2

The key to this result is a novel anytime confidence bound for sums of bounded random variables, which requires a significant departure from previous analyses of KL-based confidence bounds. The practical benefit of lil-KLUCB is illustrated in terms of the New Yorker Caption Contest problem [10]. The goal of that crowdsourcing task is to identify the funniest cartoon caption from a batch of n ? 5000 captions submitted to the contest each week. The crowd provides ?3-star? ratings for the captions, which can be mapped to {0, 1/2, 1}, for example. Unfortunately, many of the captions are not funny, getting average ratings close to 0 (and consequently very small variances). This fact, however, is ideal for KL-based confidence intervals, which are significantly tighter than those based on sub-Gaussianity and the worst-case variance of 1/4. Compared to existing methods, the lil-KLUCB algorithm better addresses the two key features in this sort of application: (1) a very large number of arms, and (2) bounded reward distributions which, in many cases, have very low variance. In certain instances, this can have a profound effect on sample complexity (e.g., O(n2) complexity for algorithms using sub-Gaussian bounds vs. O(n log n) for lil-KLUCB, as shown in Table 1). The paper is organized as follows. Section 2 defines the best-arm identification problem, gives the lil-KLUCB algorithm and states the main results. We also briefly review related literature, and compare

the performance of lil-KLUCB to that of previous algorithms. Section 3 provides the main technical contribution of the paper, a novel anytime confidence bound for sums of bounded random variables. Section 4 analyzes the performance of the lil-KLUCB algorithm. Section 5 provides experimental support for the lil-KLUCB algorithm using data from the New Yorker Caption Contest.

2

Problem Statement and Main Results

Consider a MAB problem with n arms. We use the shorthand notation [n] := {1, . . . , n}. For every i ? [n] let {Xi,j }j?N denote the reward sequence of arm i, and suppose that P(Xi,j ? [0, 1]) = 1 for all i ? [n], j ? N. Furthermore, assume that all rewards are independent, and that Xi,j ? Pi for all j ? N. Let the mean reward of arm i be denoted by ?i and assume w.l.o.g. that ?1 ¿ ?2 ? ? ? ? ? ?n . We focus on the best-arm identification problem in the fixed-confidence setting. At every time t ? N we are allowed to select an arm to sample (based on past rewards) and observe the next element in its reward sequence. Based on the observed rewards, we wish to find the arm with the highest mean reward. In the fixed confidence setting, we prescribe a probability of error ? ? (0, 1) and our goal is to construct an algorithm that finds the best arm with probability at least 1 ? ?. Among 1 ? ? accurate algorithms, one naturally favors those that require fewer samples. Hence proving upper bounds on the sample complexity of a candidate algorithm is of prime importance. The lil-KLUCB algorithm that we propose is a fusion of lil-UCB [5] and KL-LUCB [8], and its operation is essentially a special instance of LUCB++ [11]. At each time step t, let Ti (t) denote the total number of samples drawn from arm i so far, and let ? bi,Ti (t) denote corresponding empirical mean. The algorithm is based on lower and upper confidence bounds of the following general form: 4

A more precise characterization of the sample complexity is given in Section 2. The Chernoff-information between random variables Ber(x) and Ber(y) (0 ¡ x ¡ y ¡ 1) is D? (x, y) = ? 1?z D(z , x) = D(z ? , y), where D(z, x) = z log xz + (1 ? z) log 1?x and z ? is the unique z ? (x, y) such that D(z, x) = D(z, y). 5

2

for each i ? [n] and any ? ? (0, 1)

c log (? log2 (2Ti (t))/?) Li (t, ?) = inf m ¡ ? bi,Ti (t) : D ? bi,Ti (t) , m ? Ti (t)

c log (? log2 (2Ti (t))/?) Ui (t, ?) = sup m ¿ ? bi,Ti (t) : D ? bi,Ti (t) , m ? Ti (t) where c and ? are small constants (defined in the next section). These bounds are designed so that with probability at least 1 ? ?, Li (Ti (t), ?) ? ?i ? Ui (Ti (t), ?) holds for all t ? N. For any t ? N let TOP(t) be the index of the arm with the highest empirical mean, breaking ties at random. With this notation, we state the lil-KLUCB algorithm and our main theoretical result. lil-KLUCB 1. Initialize by sampling every arm once. 2. While LTOP(t) (TTOP(t) (t), ?/(n ? 1)) ?

max Ui (Ti (t), ?) do:

i6=TOP(t)

? Sample the following two arms: ? TOP(t), and ? arg max Ui (Ti (t), ?) i6=TOP(t)

and update means and confidence bounds. 3. Output TOP(t) Theorem 1. For every i ? 2 let ? ei ? (?i , ?1), and ? e = maxi?2 ? ei . With probability at least 1 ? 2?, lil-KLUCB returns the arm with the largest mean and the total number of samples it collects is upper bounded by

X c0 log ? ?1 log D? (?i , ? ei)?1 c0 log (n ? 1)? ?1 log D? (?1 , ? e)?1 inf + , ? e2 ,...,e ?n D? (?1 , ? e) D? (?i , ? ei) i?2

?

where c0 is some universal constant, D (x, y) is the Chernoff-information. Remark 1. Note that the LUCB++ algorithm of [11] is general enough to handle identification of the top k arms (not just the best-arm). All arguments presented in this paper also go through when considering the top-k problem for k ¿ 1. However, to keep the arguments clear and concise, we chose to focus on the best-arm problem only. 2.1

Comparison with previous work

We now compare the sample complexity of lil-KLUCB to that of the two most closely related algorithms, KL-LUCB [8] and lil-UCB [5]. For a detailed review of the history of MAB problems and the use of KL-confidence intervals for bounded rewards, we refer the reader to [3, 9, 4]. For the KL-LUCB algorithm, Theorem 3 of [8] guarantees a high-probability sample complexity upper bound scaling as X

inf (D? (?i , c))?1 log n? ?1 (D? (?i , c))?1 . c?(?1 ,?2)

i?1

Our result improves this in two ways. On one hand, we eliminate the unnecessary logarithmic dependence on the number of arms n in every term. Note that the log n factor still appears in Theorem 1 in the term corresponding to the number of samples on the best arm. It is shown in [11] that this factor is indeed unavoidable. The other improvement lil-KLUCB offers over KL-LUCB is improved logarithmic dependence on the Chernoff-information terms. This is due to the tighter confidence intervals derived in Section 3. Comparing Theorem 1 to the sample complexity of lil-UCB, we see that the two are of the same form, the exception being that the Chernoff-information terms take the place of the squared mean-gaps 3

(which arise due to the use of sub-Gaussian (SG) bounds). To give a sense of the improvement this can provide, we compare the sums6 X X 1 1 SKL = . and SSG = ? D (?i , ?1) ?2i i?2

0

0

i?2

0

Let ?, ? ? (0, 1), ? ¡ ? and ? = —? ? ? —. Note that the Chernoff-information between Ber(?) and Ber(?0) can be expressed as D? (?, ?0) = max 0 min{D(x, ?), D(x, ?0)} = D(x? , ?) = D(x? , ?0) = x?[?,?]

D(x? ,?)+D(x? ,?0) 2

,

for some unique x? ? [?, ?0]. It follows that D? (?, ?0) ?

D(x, ?) + D(x, ?0) 1 p . = log p 2 x?[?,?] ?(? + ?) + (1 ? ?)(1 ? ? ? ?) min 0

Using this with every term in SKL gives us an upper bound on that sum. If the means are all bounded well away from 0 and 1, then SKL may not differ that much from SSG . There are some situations however, when the two expressions behave radically differently. As an example, consider a situation when ?1 = 1. In this case we get X X 1 X 1 2 ? 2 SKL ? = SSG .

1 ?i ?2i log 1??i i?2

i?2

i?2

Table 1 illustrates the difference between the scaling of the sums SKL and SSG when the gaps have the parametric form ?i = (i/n)? . Table 1: SKL versus SSG for mean gaps ?i = (ni)? , i = 1, . . . , n ? SKL SSG

? (0, 1/2)

1/2

? (1/2, 1)

1

? (1, ?)

n n

n n log n

n n2?

n log n n2

n? n2?

We see that KL-type confidence bounds can sometimes provide a significant advantage in terms of the sample complexity. Intuitively, the gains will be greatest when many of the means are close to 0 or 1 (and hence have low variance). We will illustrate in Section 5 that such gains often also manifest in practical applications like the New Yorker Caption Contest problem.

3

Anytime Confidence Intervals for Sums of Bounded Random Variables

The main step in our analysis is proving a sharp anytime confidence bound for the mean of bounded random variables. These will be used to show, in Section 4, that lil-KLUCB draws at most O((Di?)?1 log log(Di?)?1) samples from a suboptimal arm i, where Di? := D? (?1 , ?i) is the Chernoff-information between a Ber(?1) and a Ber(?i) random variable and arm 1 is the arm with the largest mean. The iterated log factor is a necessary consequence of the law-of-the-iterated logarithm [5], and in it is in this sense that we call the bound sharp. Prior work on MAB algorithms based on KL-type confidence bounds [4, 9, 3] did not focus on deriving tight anytime confidence bounds. Consider a sequence of iid random variables Y1 , Y2 , . . . that are bounded in [0, 1] and have mean ?. P Let ? bt = 1t j?[t] Yj be the empirical mean of the observations up to time t ? N. Theorem 2. Let ? ? [0, 1] and ? ? (0, 1) be arbitrary. Fix any l ? 0 and set N = 2l , and define ? NN+1 ? X X N +1 N +1 ?(N) = ? 1/(N +1) ? 1{l6=0} log2 (2t)? N + N (k + 1)? N ? . k?l

t?[N] 6

Consulting the proof of Theorem 1 it is clear that the number of samples on the sub-optimal arms of lil-KLUCB scales essentially as SKL w.h.p. (ignoring doubly logarithmic terms), and a similar argument can be made about lil-UCB. This justifies considering these sums in order to compare lil-KLUCB and lil-UCB.

4

(i) Define the sequence zt ? (0, 1 ? ?], t ? N such that log (?(N) log (2t)/?) 2 D ? + NN+1 zt , ? = , t if a solution exists, and zt = 1 ? ? otherwise. Then P (?t ? N : ? bt ? ? ¿ zt) ? ?.

(1)

(ii) Define the sequence zt ¿ 0, t ? N such that

log (?(N) log (2t)/?) 2 D ? ? NN+1 zt , ? = , t if a solution exists, and zt = ? otherwise. Then P (?t ? N : ? bt ? ? ¡ ?zt) ? ?. The result above can be used to construct anytime confidence bounds for the mean as follows. Consider part (i) of Theorem 2 and fix ?. The result gives a sequence zt that upper bounds the deviations of the empirical mean. It is defined through an equation of the form D(? + N zt /(N + 1), ?) = ft . Note that the arguments of the function on the left must be in the interval [0, 1], in particular N zt /(N + 1) ¡ 1 ? ?, and the maximum of D(? + x, ?) for x ¿ 0 is D(1, ?) = log ??1 . Hence, equation 1 does not have a solution if ft is too large (that is, if t is small). In these cases we set zt = 1 ? ?. However, since ft is decreasing, equation 1 does have a solution when t ? T (for some T depending on ?), and this solution is unique (since D(? + x, ?) is strictly increasing). With high probability ? bt ? ? ? zt for all t ? N by Theorem 2. Furthermore, the function D(? + x, ?) is increasing in x ? 0. By combining these facts we get that with probability at least 1 ? ?

t +? D ? + NN+1 zt , ? ? D NN?b+1 ,? . On the other hand

log (?(N) log2 (2t)/?) , t by definition. Chaining these two inequalities leads to the lower confidence bound

log (?(N) log (2t)/?) N? bt +m 2 L(t, ?) = inf m ¡ ? bt : D N +1 , m ? t D ?+

N N +1 zt , ?

?

(2)

which holds for all times t with probability at least 1 ? ?. Considering the left deviations of ? bt ? ? we can get an upper confidence bound in a similar manner:

log (?(N) log (2t)/?) +m 2 U (t, ?) = sup m ¿ ? bt : D NN?bt+1 ,m ? . (3) t That is, for all times t, with probability at least 1 ? 2? we have L(t, ?) ? ? bt ? U (t, ?). Note that the constant log ?(N) ? 2 log2 (N), so the choice of N plays a relatively mild role in +m the bounds. However, we note here that if N is sufficiently large, then NN?bt+1 ? ? bt , and thus

N? bt +m D N +1 , m ? D (b ?t , m), in which case the bounds above are easily compared to those in prior works [4, 9, 3]. We make this connection more precise and show that the confidence intervals defined as

c(N) log (?(N) log2 (2t)/?) L0 (t, ?) = inf m ¡ ? bt : D (b ?t , m) ? , and t

c(N) log (?(N) log2 (2t)/?) U 0 (t, ?) = inf m ¿ ? bt : D (b ?t , m) ? , t satisfy L0 (t, ?) ? ? bt ? U 0 (t, ?) for all t, with probability 1 ? 2?. The constant c(N) is defined in Theorem 1 in the Supplementary Material, where the correctness of L0 (t, ?) and U 0 (t, ?) is shown. Proof of Theorem 2. The proofs of parts (i) and (ii) are completely analogous, henceP in what follows we only prove part (i). Note that {b ?t ? ? ¿ zt } ?? {St ¿ tzt }, where St = j?[t] (Yj ? ?) denotes the centered sum up to time t. We start with a simple union bound X

P ?t ? [2k , 2k+1] : St ¿ tzt . (4) P (?t ? N : St ¿ tzt) ? P (?t ? [N] : St ¿ tzt) + k?l

5

First, we bound each summand in the second term individually. In an effort to save space, we define the event Ak = {?t ? [2k , 2k+1] : St ¿ tzt }. Let tj,k = (1 + Nj)2k . In what follows we use the notation tj ? tj,k . We have X X

P (Ak) ? P (?t ? [tj?1 , tj] : St ¿ tzt) ? P ?t ? [tj?1 , tj] : St ¿ tj?1 ztj?1 , j?[N]

j?[N]

where the last step is true if tzt is non-decreasing in t. This technical claim is formally shown in Lemma 1 in the Supplementary Material. However, to give a short heuristic, it is easy to see that tzt has an increasing lower bound. Noting that D(? + x, ?) is convex in x (the second derivative is positive), and that D(?, ?) = 0, we have D(1, ?)x ? D(? + x, ?). Hence zt & t?1 log log t. Using a Chernoff-type bound together with Doob?s inequality, we can continue as X

P (Ak) ? inf P ?t ? [tj?1 , tj] : exp (?St) ¿ exp ?tj?1 ztj?1 ?¿0

?

j?[N]

?Stj exp ? sup ?tj?1 ztj?1 ? log E e

X

?¿0

j?[N]

=

X

exp ?tj sup ??0

j?[N]

? NN+j?1 +j ztj?1

? log E e

?(Y1 ??)

.

Using E(e?Y1) ? E(e??) where ? ? Ber(?) (see Lemma 9 of [4]), and the notation ?j =

X P (Ak) ? exp ?tj sup ??j ztj?1 ? log E e?(???) =

N +j?1 N +j ,

??0

j?[N]

X

(5)

exp ?tj D ? + ?j ztj?1 , ? ,

(6)

j?[N]

since the rate function of a Bernoulli random variable can be explicitly computed, namely we have sup?¿0 (?x ? log E(e??)) = D(? + x, ?) (see [2]). Again, we use the convexity of D(? + x, ?). For any ? ? (0, 1) we have ?D(? + x, ?) ? N D(? + ?x, ?), since D(?, ?) = 0. Using this with ? = ?j (N +1) and x = ?j ztj?1 , we get that

N N ?j (N +1) D ? + ?j ztj?1 , ? ? D ? + N +1 ztj?1 , ? . This implies P (Ak) ?

X

exp ?tj NN+1 ?j D ? +

N N +1 ztj?1 , ?

.

(7)

j?[N]

Plugging in the definition of tj and the sequence zt , and noting that ? ¡ 1, we arrive at the bound

NN+1 X N +1 N +j?1 ? log ?(N) log2 (2k+1)/? ? N ?(N)(k+1) P (Ak) ? exp ? . N N j?[N]

Regarding the first term in (4), again using the Bernoulli rate function bound we have X X P (?t ? [N] : ? bt ? ? ¿ zt) ? P (b ?t ? ? ¿ zt) ? exp (?tD(? + zt , ?)) . t?[N]

t?[N]

Using the convexity of D(? + x, ?) as before, we can continue as

X P (?t ? [N] : ? bt ? ? ¿ zt) ? exp ?t NN+1 D ? +

N N +1 zt , ?

t?[N]

?

X

exp ? NN+1 log (?(N) log2 (2t)/?)

t?[N]

??

N +1 N

?(N)?

N +1 N

X t?[N]

6

log2 (2t)?

N +1 N

.

Plugging the two bounds back into (4) we conclude that ? ? X N +1 N +1 X N +1 N +1 ?1{l6=0} log2 (2j)? N + (k + 1)? N ? ? ? , P (?t : ? bt ? ? ¿ zt) ? ? N ?(N)? N k?l

j?[N]

by the definition of ?(N).

4

Analysis of lil-KLUCB

Recall that the lil-KLUCB algorithm uses confidence bounds of the form Ui (t, ?) = sup{m ¿ ? bt : D(b ?t , m) ? ft (?)} with some decreasing sequence ft (?). In this section we make this dependence explicit, and use the notations Ui (ft (?)) and Li (ft (?)) for upper and lower confidence bounds. For any ¿ 0 and i ? [n], define the events ?i () = {?t ? N : ?i ? [Li (ft ()), Ui (ft ())]}. The correctness of the algorithm follows from the correctness of the individual confidence intervals, as is usually the case with LUCB algorithms. This is shown formally in Proposition 1 provided in the Supplementary Materials. The main focus in this section is to show a high probability upper bound on the sample complexity. This can be done by combining arguments frequently used for analyzing LUCB algorithms and those used in the analysis of the lil-UCB [5]. The proof is very similar in spirit to that of the LUCB++ algorithm [11]. Due to spatial restrictions, we only provide a proof sketch here, while the detailed proof is provided in the Supplementary Materials. Proof sketch of Theorem 1. Observe that at each time step two things can happen (apart from stopping): (1) Arm 1 is not sampled (two sub-optimal arms are sampled); (2) Arm 1 is sampled together with some other (suboptimal) arm. Our aim is to upper bound the number of times any given arm is sampled for either of the reasons above. We do so by conditioning on the event ? ?

?0 = ?1 (?) ? ? ?i (?i)? , for a certain choice of {?i } defined below. i?2

For instance, if arm 1 is not sampled at a given time t, we know that TOP(t) 6= 1, which means there must be an arm i ? 2 such that Ui (Ti (t), ?) ? U1 (T1 (t), ?). However, on the event ?1 (?), the UCB of arm 1 is accurate, implying that Ui (Ti (t), ?) ? ?1 . This implies that Ti (t) can not be too big, since on ?i (?i), ? bi,t is ?close" to ?i , and also Ui (Ti (t), ?) is not much larger then ? bi . All this is made formal in Lemma 2, yielding the following upper bound on number of times arm i is sampled for reason (1): ?i (? ? ?i) = min {t ? N : ft (? ? ?i) ¡ D? (?i , ?1)} . Similar arguments can be made about the number of samples of any suboptimal arm i for reason (2), and also the number of samples on arm 1. This results in the sample complexity upper bound

X K1 log ? ?1 log D? (?i , ? ei)?1 + log ?i?1 K1 log (n ? 1)? ?1 log D? (?1 , ? e)?1 + , D? (?1 , ? e) D? (?i , ? ei) i?2

0

on the event ? , where K1 is a universal constant. Finally, we define the quantities ?i = sup{ ¿ 0 : Ui (ft ()) ? ?i ?t ? N}. Note that we have P(?i ¡ ?) = P(?t ? N : Ui (ft (?)) ? ?i) ? ? according to Theorem 1 in the Supplementary Material. Substituting ? = exp(?D? (?i , ? ei)z) we get

log ?i?1 ei)z) . P D? (?i ,e?i) ? z ? exp(?D? (?i , ? Hence {?i }i?2 are independent sub-exponential variables, which allows us to control their contribution to the sum above using standard techniques.

5

Real-World Crowdsourcing

We now compare the performance of lil-KLUCB to that of other algorithms in the literature. We do this using both synthetic data and real data from the

New Yorker Cartoon Caption contest [10]7 . To 7

These data can be found at https://github.com/nextml/caption-contest-data

7

keep comparisons fair, we run the same UCB algorithm for all the competing confidence bounds. We set N = 8 and ? = 0.01 in our experiments. The confidence bounds are [KL]: the KL-bound derived based on Theorem 2, [SG1]: a matching sub-Gaussian bound derived using the proof of Theorem 2, using sub-Gaussian tails instead of the KL rate-function (the exact derivations are in the Supplementary Material), and [SG2]: the sharper sub-Gaussian bound provided by Theorem 8 of [7].

0.00

0.75

1.50

2.25

3.00

3.75

1.0 0.8 0.6 0.4 0.2

0.6 0.4 0.2

Kaufmann lil?UCB KL?UCB SG lil?UCB

1.2

? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?

? ?
? ?

(Empirical) probability ? 250 trials

? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ? ? ? ?

Kaufmann lil?UCB KL?UCB SG lil?UCB

0.0

1.2

P(best arm in top 5), alpha=1/2

0.8

1.0

P(best arm in top 5), alpha=1

0.0

(Empirical) probability ? 250 trials

We compare these methods by computing the empirical probability that the best-arm is among the top 5 empirically best arms, as a function of the total number of samples. We do so using using synthetic data in Figure 5 , where the Bernoulli rewards simulate cases from Table 1, and using real human response data from two representative New Yorker caption contests in Figure 5.

4.50

0.00

Number of samples (10 thousands)
0.15
0.30
0.45
0.60
0.75
0.90
Number of samples (10 thousands)
Figure 1: Probability of the best-arm in the top 5 empirically best arms, as a function of the number of samples, based on 250 repetitions. ?i = 1 ? ((i ? 1)/n)? , with ? = 1 in the left panel, and ? = 1/2 in the right panel. The mean-profile is shown above each plot. [KL] Blue; [SG1] Red; [SG2] Black.

As seen in Table 1, the KL confidence bounds have the potential to greatly outperform the subGaussian ones. To illustrate this indeed translates into superior performance, we simulate two cases, with means ?i = 1 ? ((i ? 1)/n)? , with ? = 1/2 and ? = 1, and n = 1000. As expected, the KL-based method requires significantly fewer samples (about 20 % for ? = 1 and 30 % for ? = 1/2) to find the best arm. Furthermore, the arms with means below the median are sampled about 15 and 25 % of the time respectively ? key in crowdsourcing applications, since having participants answer fewer irrelevant (and potentially annoying) questions improves both efficiency and user experience.

0
2
4
6
8
10
12
14
16
18
20
1.2 1.0 0.8 0.6 0.4 0.2
Kaufmann lil?UCB KL?UCB SG lil?UCB
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?

? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?

Kaufmann lil?UCB KL?UCB SG lil?UCB

0.0

0.2

0.4

0.6

0.8

1.0

1.2

(Empirical) probability ? 250 trials

P(best arm in top 5), Contest 512

? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?

? ?

0.0

(Empirical) probability ? 250 trials

P(best arm in top 5), Contest 558

0.00

Number of samples (10 thousands)

0.75

1.50

2.25

3.00

3.75

4.50

Number of samples (10 thousands)

Figure 2: Probability of the best-arm in the top 5 empirically best arms vs. number of samples, based on 250 bootstrapped repetitions. Data from New Yorker contest 558 (?1 = 0.536) on left, and contest 512 (?1 = 0.8) on right. Mean-profile above each plot. [KL] Blue; [SG1] Red; [SG2] Black.

8

To see how these methods fair on real data, we also run these algorithms on bootstrapped human response data from the real New Yorker Caption Contest. The mean reward of the best arm in these contests is usually between 0.5 and 0.85, hence we choose one contest from each end of this spectrum. At the lower end of the spectrum, the three methods fair comparably. This is expected because the sub-Gaussian bounds are relatively good for means about 0.5. However, in cases where the top mean is significantly larger than 0.5 we see a marked improvement in the KL-based algorithm.

Extension to numerical experiments Since a large number of algorithms have been proposed in the literature for best arm identification, we include another algorithm in the numerical experiments for comparison. Previously we compared lil-KLUCB to lil-UCB as a comparison for two reasons. First, this comparison illustrates best the gains of using the novel anytime confidence bounds as opposed to those using sub-Gaussian tails. Second, since lil-UCB is the state of the art algorithm, any other algorithm will likely perform worse. The authors of [6] compare a number of different best arm identification methods, and conclude that two of them seem to stand out: lil-UCB and Thompson sampling. Therefore, we now include Thomspon sampling [Th] in our numerical experiments for the New Yorker data. We implemented the method as prescribed in [6]. As can bee seen in Figure 5, Thompson sampling seems to perform somewhat worse than the previous methods in these two instances.

0

2

4

6
8
10
12
14
16
18
20
1.2 1.0 0.8 0.6 0.4 0.2
Kaufmann lil?UCB KL?UCB SG lil?UCB Thompson
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ?
? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?
Kaufmann lil?UCB KL?UCB SG lil?UCB Thompson
0.0
0.2
0.4
0.6
0.8
1.0
1.2
(Empirical) probability ? 250 trials
P(best arm in top 5), Contest 512
? ?
? ?

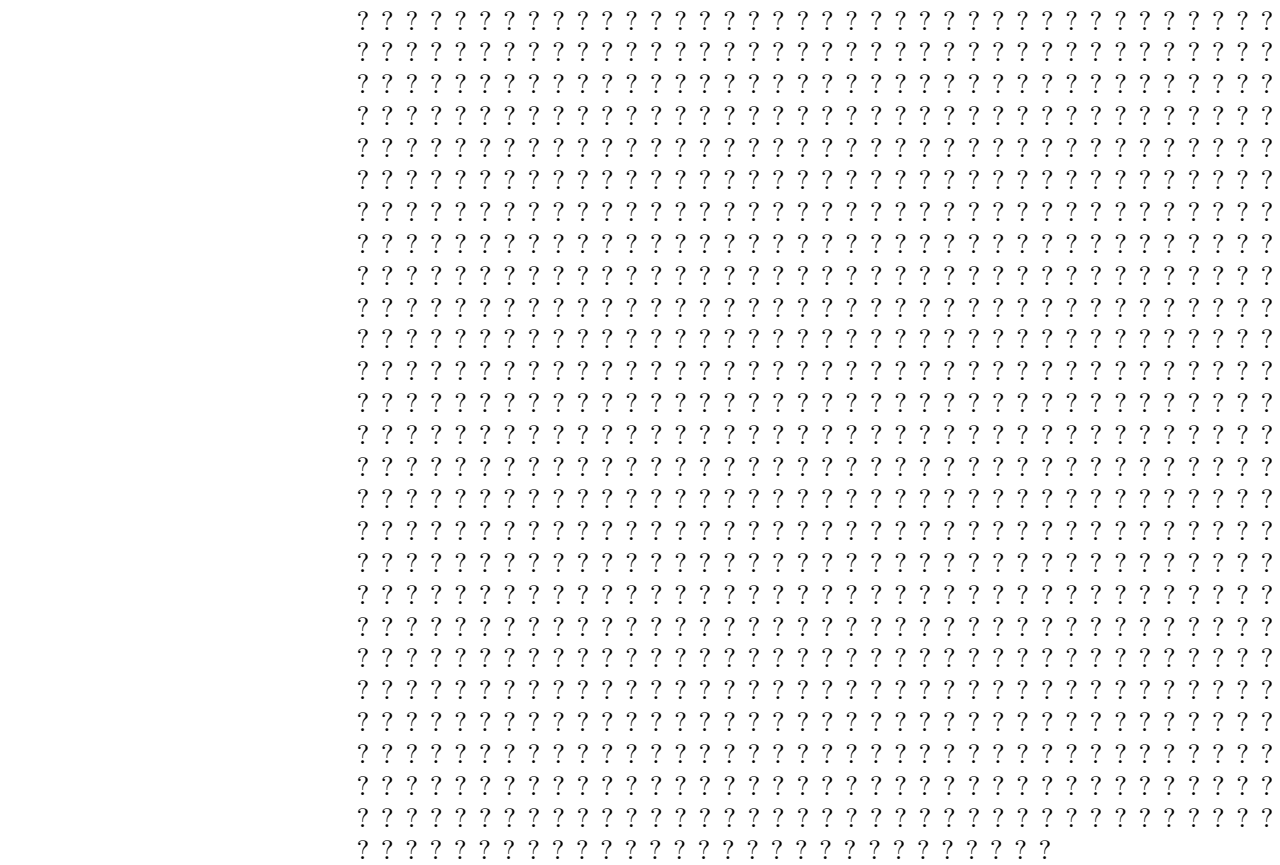

Figure 3: Probability of the best-arm in the top 5 empirically best arms vs. number of samples, based on 250 bootstrapped repetitions. Data from New Yorker contest 558 (?1 = 0.536) on left, and contest 512 (?1 = 0.8) on right. Mean-profile above each plot. [KL] Blue; [SG1] Red; [SG2] Black; [Th] Purple.

9

2 References

[1] Jean-Yves Audibert and S?bastien Bubeck. Best arm identification in multi-armed bandits. In COLT-23th Conference on Learning Theory-2010, pages 13?p, 2010. [2] St?phane Boucheron, G?bor Lugosi, and Pascal Massart. Concentration inequalities: A nonasymptotic theory of independence. Oxford university press, 2013. [3] Olivier Capp?, Aur?lien Garivier, Odalric-Ambrym Maillard, R?mi Munos, Gilles Stoltz, et al. Kullback? leibler upper confidence bounds for optimal sequential allocation. The Annals of Statistics, 41(3):1516?1541, 2013. [4] Aur?lien Garivier and Olivier Capp?. The kl-ucb algorithm for bounded stochastic bandits and beyond. In COLT, pages 359?376, 2011. [5] Kevin Jamieson, Matthew Malloy, Robert Nowak, and S?bastien Bubeck. lil-ucb: An optimal exploration algorithm for multi-armed bandits. In Conference on Learning Theory, pages 423?439, 2014. [6] Kevin G Jamieson, Lalit Jain, Chris Fernandez, Nicholas J Glattard, and Rob Nowak. Next: A system for real-world development, evaluation, and application of active learning. In Advances in Neural Information Processing Systems, pages 2656?2664, 2015. [7] Emilie Kaufmann, Olivier Capp?, and Aur?lien Garivier. On the complexity of best arm identification in multi-armed bandit models. The Journal of Machine Learning Research, 2016. [8] Emilie Kaufmann and Shivaram Kalyanakrishnan. Information complexity in bandit subset selection. In COLT, pages 228?251, 2013. [9] Odalric-Ambrym Maillard, R?mi Munos, Gilles Stoltz, et al. A finite-time analysis of multi-armed bandits problems with kullback-leibler divergences. In COLT, pages 497?514, 2011. [10] B. Fox Rubin. How new yorker cartoons could teach computers to be funny. CNET News, 2016. https://www.cnet.com/news/how-new-yorker-cartoons-could-teach-computers-to-be-funny/. [11] Max Simchowitz, Kevin Jamieson, and Benjamin Recht. The simulator: Understanding adaptive sampling in the moderate-confidence regime. arXiv preprint arXiv:1702.05186, 2017.

10