An Engineered Empirical Bernstein Bound*

Mark A. Burgess[1] (✉), Archie C. Chapman[2], and Paul Scott[1]

[1] Australian National University,
College of Engineering & Computer Science, ACT, 0200, Australia.
markburgess1989@gmail.com, mark.burgess@anu.edu.au
paul.scott@anu.edu.au

[2] University of Sydney,
School of Electrical and Information Engineering, NSW, 2006, Australia
archie.chapman@sydney.edu.au

Abstract. We derive a tightened *empirical Bernstein bound* (EBB) on the variation of the sample mean from the population mean, and show that it improves the performance of *upper confidence bound* (UCB) methods in multi-armed bandit problems. Like other EBBs, our EBB is a concentration inequality for the variation of the sample mean in terms of the sample variance. Its derivation uses a combination of probability unions and Chernoff bounds for the mean of samples and mean of sample squares. Analysis reveals that our approach can tighten the best existing EBBs by about a third, and thereby halves the distance to a bound constructed with perfect variance information. We illustrate the practical usefulness of our novel EBB by applying it to a multi-armed bandit problem as a component of a UCB method. Our method outperforms existing approaches by producing lower expected regret than variants of UCB employing several other bounds, including state-of-the-art EBBs.

Keywords: Concentration inequality · Chernoff bounds · Hoeffding's inequality · Empirical Bernstein bound

1 Introduction

Data-driven processes and decision-making applications typically rely on sample statistics to infer parameters of a population or evaluate decision options. Depending on the domain, different assumptions can be made about the distribution of the data, which in turn determine which computational routines are used to compute the required population statistics. These assumptions may be based on prior information, expert opinion, or determined from the characteristics of the system under observation.

Within this context, finite-sample *concentration inequalities* are used to place bounds on the variation of sample statistics around their population values. Such bounds are applied in a range of data science contexts for a variety of prediction, machine learning and hypothesis testing tasks, including: change detection [19,

* A great thanks to Sylvie Thiébaux for advice and encouragement.

11] and classification [25] in data streams; outlier analysis in large databases [2]; online optimisation [17, 1]; and, of most relevance to this paper, online prediction and learning problems [15, 23, 30, 22], particularly in settings with *bandit feedback* [5, 3, 31]. In particular, the recently developed *empirical Bernstein bounds* (EBB) are of significant interest [22, 4]. These are probability bounds describing the likely difference of a sample mean from the population mean in terms of the *sample* variance, under the assumption that the population data is bounded within an interval of known width. EBBs have been used as a method of generating confidence bounds for the mean, and an outstanding task is to see how much these techniques can be improved.

Given this challenge, in this work, we take inspiration and extend the work of Maurer and Pontil ([22], M&P in the remainder) to develop a new EBB. Our EBB tightens existing bounds by incorporating a combination of bounds on the variation of the sample variance. Specifically, we use two Chernoff bounds, for the sample mean and the mean of sample squares, which are fused using a probability union and variance decomposition, to create a novel probability bound for the sample variance, which is then used to derive our novel EBB.

Evaluations show that our EBB significantly tightens the current state-of-the-art bounds. Specifically, our EBB can shrink the best existing EBBs by about a third. This represents half of the distance between the best existing EBBs and an unattainable Bernstein bound constructed with perfect variance information. Moreover, we demonstrate the use of our novel EBB in an *upper-confidence bound* (UCB) multi-armed bandit (MAB) algorithm. Results from a set of MABs show that using our bound in a UCB algorithm outperforms existing approaches, by producing comparable or lower expected regret than employing other existing bounds, including state-of-the-art EBBs.

The paper is organised as follows. Related work and preliminary concepts are reviewed in Sections 2 and 3, respectively. Our main results are in Section 4, where we derive a novel EBB. In Section 5 we evaluate our EBB and show its improvements over existing bounds. In Section 6 we apply it to a multi-armed bandit problem as part of a UCB algorithm, which demonstrates how our tighter EBB improves the algorithm's learning performance. Section 7 concludes.

2 Related Work

Concentration inequalities are probabilistic bounds describing how far a random variable is expected to deviate from (or otherwise be concentrated around) a particular value. Most classic concentration inequalities describe the expected deviation of sample statistics, including Chebyshev's inequality [12], the Bernstein's inequalities [10], Hoeffding's inequalities [18] and Bennett's inequalities [7]. Building on these, new analysis has yielded a wide range concentration inequalities and methods of generating them [9, 13]. In particular, recent innovations concern the concentration of more-general functions of random variables, such as the Efron-Stein [16] and entropy methods [14], and applications of Talagrand's concentration inequality [28]. Inequalities such as these are used to describe the

expected variability of sample statistics, such as the distance of a sample mean from the population mean.

Furthermore, additional sample statistics can be used to tighten such bounds, because these statistics provide extra distributional information that are incorporated as a factor into classical inequalities. EBBs [22, 4] are one example of this, where sample variance information is used to tighten a classical Bernstein bound. However, it remains to be seen how far bounds derived by this approach can be tightened.

3 Preliminaries

To begin, we state three lemmas which form the basis for our derivation (proofs in Appendix A.1). The first is an often used result related to union bounds:

Lemma 1 (Probability Union). *For any random variables a, b and c:*

$$\mathbb{P}(a > c) \;\leq\; \mathbb{P}(a > b) + \mathbb{P}(b > c)$$

This result is used to bound the probability relationship between two variables via knowledge of the probability relationship between them and a third variable. The second definition relates the value of the sample mean and the value of sample squares to the sample variance. It is expanded here because we will later use these relationships to create bounds for the sample variance from bounds on the sample squares and sample mean.

Lemma 2 (Variance Decomposition). *For n samples x_i, sample mean $\hat{\mu} = \frac{1}{n}\sum_i x_i$, sample variance $\hat{\sigma}^2 = \frac{1}{n-1}\sum_i (x_i - \hat{\mu})^2$, and average of sample squares $\hat{\sigma}_0^2 = \frac{1}{n}\sum_i x_i^2$, the following relationship holds:*

$$\hat{\sigma}_0^2 = \hat{\mu}^2 + \frac{n-1}{n}\hat{\sigma}^2$$

In order to derive our novel bound, we use the next lemma, which encapsulates a range of inequalities called *Chernoff bounds* that give bounds on the mean of random variables:

Lemma 3 (Chernoff Bound). *If $\hat{\mu}$ is sample mean of n independent and identically distributed samples of random variable X then for any $s > 0$ and t:*

$$\mathbb{P}(\hat{\mu} \geq t) \leq \mathbb{E}\left[\exp(sX)\right]^n \exp(-snt)$$

The proof of this statement is straightforward and uses Markov's inequality and the i.i.d of the samples. In the next section, we use these components to derive the bounds on the sample mean and the mean of sample squares, which we then use to create a new EBB.

4 Derivation and numerical implementation

In this section, we derive two Chernoff bounds, for the sample mean and the mean of sample squares, (Lemmas 5 and 6, respectively). These are fused using a probability union and variance decomposition, defined above, to derive a bound for the sample variance. This bound is then used to derive our new EBB, as presented in Theorem 7. However, due to its analytic intractability, we complete the derivation by discussing how to numerically implement the bound.

4.1 Derivation

Our first probability bound is a Chernoff bound on the sample mean called *Bennett's inequality*. This bound is not new and was derived by [18] and [7] and has subsequently been a subject of discussion and many further developments [8, 24, 29]; we provide a proof in Appendix A.2.

Theorem 4 (Bennett's inequality). *Let X be a real-valued random variable with a mean of zero and variance σ^2, that is bounded $a \leq X \leq b$. Then for $t > 0$, the mean $\hat{\mu}$ of n samples of X is probability bounded by:*

$$\mathbb{P}(\hat{\mu} \geq t) \leq H_1^n\left(\frac{\sigma^2}{b^2}, \frac{t}{b}\right), \tag{1}$$

where:

$$H_1^n\left(\frac{\sigma^2}{b^2}, \frac{t}{b}\right) = \left(\left(\frac{\frac{\sigma^2}{b^2}}{\frac{\sigma^2}{b^2} + \frac{t}{b}}\right)^{\frac{\sigma^2}{b^2}+\frac{t}{b}} \left(1 - \frac{t}{b}\right)^{\frac{t}{b}-1}\right)^{\frac{n}{\frac{\sigma^2}{b^2}+1}}$$

We will also use a double-sided version of this bound:

$$\mathbb{P}(\hat{\mu}^2 \geq r^2) \leq H_1^n\left(\frac{\sigma^2}{b^2}, \frac{r}{b}\right) + H_1^n\left(\frac{\sigma^2}{a^2}, \frac{-r}{a}\right) \tag{2}$$

The assumption that the mean is zero can be used without a loss of generality. In this way, Bennett's inequality gives us a probability bound for the difference of the sample mean from the true mean *given the variance*.

However, often in practice the variance is unknown, but can only estimate it via a sample variance statistic. Thus, we derive a bound the difference of the sample variance from the variance as follows (proof in Appendix A.3):

Lemma 5 (Sample square bound). *Let X be a real-valued random variable with a mean of zero and variance σ^2, that is bounded $a \leq X \leq b$, if $d = \max(b, -a)$ then for $y > 0$, the mean of sample squares $\hat{\sigma}_0^2 = \frac{1}{n}\sum_i x_i^2$ is probability bounded:*

$$\mathbb{P}(\sigma^2 - \hat{\sigma}_0^2 > y) \leq H_2^n\left(\frac{\sigma^2}{d^2}, \frac{y}{d^2}\right), \tag{3}$$

where:

$$H_2^n\left(\frac{\sigma^2}{d^2}, \frac{y}{d^2}\right) = \left(\left(\frac{1-\frac{\sigma^2}{d^2}}{1+\frac{y}{d^2}-\frac{\sigma^2}{d^2}}\right)^{1+\frac{y}{d^2}-\frac{\sigma^2}{d^2}}\left(\frac{\frac{\sigma^2}{d^2}}{\frac{\sigma^2}{d^2}-\frac{y}{d^2}}\right)^{\frac{\sigma^2}{d^2}-\frac{y}{d^2}}\right)^n$$

It is worth noting that we choose to restrict the use of function H_2^n to cases which are sensible for it to be applied: (i) it is defined for $a < 0 < b$, because otherwise the mean could not be zero), and (ii) $\sigma^2 \leq -ab \leq (b-a)^2/4$ by Popoviciu's inequality [27], as it is not possible for the variance to be larger given the width of the data bounds. It is important that these domain restrictions are conserved with the analysis.

At this point, we have a probability bound on the mean squared (Equation 2) and a probability bound on the sample squares (Lemma 5). With these in hand, we use lemma 2 to create a bound on the sample variance, as follows.

Lemma 6 (Sample Variance Bound). *For a random variable that is bounded $a \leq X \leq b$ with variance σ^2 and a mean of zero, if $d = \max(b, -a)$ then for $w > 0$, the sample variance $\hat{\sigma}^2$ of n samples is probability bounded by:*

$$\mathbb{P}(\sigma^2 - \hat{\sigma}^2 > w) \leq H_3^n(a, b, w, \sigma^2), \tag{4}$$

where:

$$H_3^n(a,b,w,\sigma^2) = \min_{\phi\in[0,1]}\left\{\begin{array}{l} H_1^n\left(\frac{\sigma^2}{b^2}, \frac{\sqrt{\phi(\frac{n-1}{n}w+\frac{1}{n}\sigma^2)}}{b}\right) \\ +H_1^n\left(\frac{\sigma^2}{a^2}, \frac{-\sqrt{\phi(\frac{n-1}{n}w+\frac{1}{n}\sigma^2)}}{a}\right) \\ +H_2^n\left(\frac{\sigma^2}{d^2}, \frac{(1-\phi)(\frac{n-1}{n}w+\frac{1}{n}\sigma^2)}{d^2}\right)\end{array}\right\}$$

A proof is provided in Appendix A.3. The use of the function H_3^n is subject to the same restrictions on its domain as H_2^n. Thus, in Lemma 4 we have a bound for the sample mean given the variance, and in Lemma 6 we have a probability bound for the difference of the sample variance from the population variance. Next, we outline a method of combining these two to create a bound for the sample mean given the sample variance — and thereby derive a new empirical Bernstein bound. To do this, we now expound a theorem that embodies a process followed by M&P [22].

Before beginning, we introduce some notation. For a function f with ordered inputs, we denote the inverse of f with respect to its ith input (counting from one) as $f^{-(i)}$, assuming it exists. Denote probability bounds on the differences of the sample mean from the mean, and the sample variance from the variance, by $\mathbb{P}(\hat{\mu} - \mu > t) \leq h(\sigma^2, t)$ and $\mathbb{P}(\sigma^2 - \hat{\sigma}^2 > w) \leq f(\sigma^2, w)$, respectively. Note that functions h and f have arguments σ^2 and t, and σ^2 and w, respectively.

Theorem 7 (Essential EBB). *Assume $f^{-(2)}$ and $h^{-(2)}$ both exist, and also if $h^{-(2)}$ is monotonically increasing in its first argument, so that we can define:*

$$z(\sigma^2, w) = \sigma^2 - f^{-(2)}\left(\sigma^2, w\right)$$

If $z^{-(1)}$ exists and is monotonic increasing in its first argument, then for any $x \in [0, y]$, the following relationship holds:

$$\mathbb{P}\left(\hat{\mu} - \mu > h^{-(2)}\left(z^{-(1)}\left(\hat{\sigma}^2, y - x\right), x\right)\right) \leq y$$

Proof. Substituting w for $f^{-2}(\sigma^2, w)$ gives:

$$\begin{aligned} w &\geq \mathbb{P}\left(\sigma^2 - \hat{\sigma}^2 > f^{-(2)}\left(\sigma^2, w\right)\right) \\ &\geq \mathbb{P}\left(z\left(\sigma^2, w\right) > \hat{\sigma}^2\right) \\ &\geq \mathbb{P}\left(\sigma^2 > z^{-(1)}\left(\hat{\sigma}^2, w\right)\right) \\ &\geq \mathbb{P}\left(h^{-2}\left(\sigma^2, t\right) > h^{-(2)}\left(z^{-(1)}\left(\hat{\sigma}^2, w\right), t\right)\right) \end{aligned}$$

Substituting t for $h^{-(2)}(\sigma^2, t)$ gives:

$$\mathbb{P}\left(\hat{\mu} - \mu > h^{-(2)}\left(\sigma^2, t\right)\right) \leq t.$$

Applying probability union (lemma 1) gives:

$$\mathbb{P}\left(\hat{\mu} - \mu > h^{-(2)}\left(z^{-(1)}\left(\hat{\sigma}^2, w\right), t\right)\right) \leq t + w.$$

Letting $y = t + w$ and $x = y - w$ completes the proof. ■

The result of this Theorem is an EBB, and our novel EBB is completed by substituting $h(\sigma^2, t) = H_1^n\left(\sigma^2/b^2, t/b\right)$ and $f(\sigma^2, w) = H_3^n\left(a, b, w, \sigma^2\right)$ into Theorem 7. Care must be taken in applying this theorem that all the assumptions hold, the inverses exist, and the domains of the functions are propagated through the analysis.

4.2 Numerical Implementation

Analytically solving this new EBB is challenging, however it is possible to evaluate it to arbitrary accuracy using numerical techniques. This section provides a high-level description of a process for calculating our EBB.[1]

This calculation is composed of three primary parts: (i) the computation of function $f(\sigma^2, w) = H_3^n(a, b, y, \sigma^2)$; (ii) verifying that the assumptions of Theorem 7 hold for $h(\sigma^2, t) = H_1$ and $f(\sigma^2, w) = H_3$, and; (iii) calculating the subsequent result of Theorem 7.

[1] sourcecode available at:
https://github.com/Markopolo141/Engineered-Empirical-Bernstein-Bound

First, the function $f(\sigma^2, w) = H_3^n(a, b, w, \sigma^2)$ is the solution to an optimization problem that solves for the minima of an objective function subject to constraint $\phi \in [0, 1]$. Despite its complexity, a solution can be found quickly using a single variable parameter sweep.

Second, it is necessary to verify the assumptions that $h^{-(2)}$, $f^{-(2)}$ and $z^{-(1)}$ exist and that $z^{-(1)}$ and $f^{-(2)}$ are monotonically increasing in their first argument. It is easy to note that $h(\sigma^2, t) = H_1^n\left(\sigma^2/b^2, t/b\right)$ is a closed-form function that is monotonically decreasing from 1 to 0 on the second argument, so $h^{-(2)}$ exists and is monotonically increasing in its first argument. However the remaining assumptions are more difficult to verify. For any function, the values that the function takes can be plotted as an array of points and the values that the inverse of that function takes can be determined by conducting coordinate swaps on those points. The values of $f(\sigma^2, w) = H_3^n(a, b, w, \sigma^2)$ were computed and were seen to be monotonically decreasing in its second argument confirming that $f^{-(2)}$ exists. The function $z(\sigma^2, w) = \sigma^2 - f^{-(2)}\left(\sigma^2, w\right)$ is then seen to be a manipulation on the coordinate swapped points of $f(\sigma^2, w) = H_3^n(a, b, w, \sigma^2)$. By coordinate swapping again, $z^{-(1)}$ was seen to be a regular function monotonically increasing on its first argument, hence satisfying assumptions.

Third, to numerically calculate the result of Theorem 7 the functions $h^{-(2)}$ and $z^{-(1)}$ were numerically evaluated by direct parameter searches and then composed as: $h^{-(2)}(z^{-(1)}(\hat{\sigma}^2, y-x), x)$ - which is the inner part of the expression of the new EBB parameterised by x explicitly and also a, b implicitly. However we typically don't know the values of a and b, but instead know the mean is somewhere within a finite interval of width $D = b - a$. Given this, we then take the worst case values of a and b consistent with a given D, and then take the best $x \in [0, y]$ subject to all other bounds.

5 Comparison to existing bounds

In this section, we make three comparisons of our results to existing concentration bounds, namely (i) Lemma 6 is compared to M&P's entropic bound, then our EBB is compared to (ii) M&P's EBB and (iii) Bennett's inequality with perfect variance information.

First, M&P's entropic bound [22] (originally presented in [21]) is given by:

$$\mathbb{P}(\sigma^2 - \hat{\sigma}^2 > w) \leq \exp\left(\frac{-(n-1)w^2}{2\sigma^2 D^2}\right) \tag{5}$$

The improvement our variance bound (Lemma 6) offers over theirs is given by:

$$Y\left(\frac{\sigma^2}{D^2}, \frac{w}{D^2}, n\right) = \exp\left(\frac{-(n-1)w^2}{2\sigma^2 D^2}\right) - \max_b H_3^n\left(D(1-b), Db, w, \sigma^2\right) \tag{6}$$

where b has a viable range between 0.5 and $0.5 - \sqrt{0.25 - \sigma^2/D^2}$ (via Popoviciu's inequality). Figure 1 plots this improvement against σ^2 and w for $n = 200$, which shows large regions of advantage. However, it is possible to use the minima

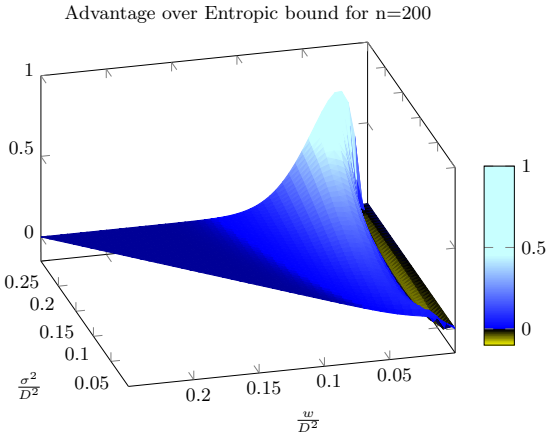

Fig. 1. The strength of our variance bound over Maurer's Entropic bound. The graph of $Y\left(\frac{\sigma^2}{D^2}, \frac{w}{D^2}, n\right)$ from Equation 6

of several different variance bounds, so in constructing our EBB, we take the minima of our variance bound and the entropic bound.

Second, we compare our EBB directly with M&P's EBB [22], given by:

$$\mathbb{P}\left(\mu - \hat{\mu} > \sqrt{\frac{2\hat{\sigma}^2 \log(2/y)}{n}} + \frac{7D\log(2/y)}{3(n-1)}\right) < y. \tag{7}$$

In order to fairly compare our EBB to M&P's we apply Popoviciu's inequality as a domain restriction, and carry it through their derivation, as we did to our own EBB. Specifically, this is the domain where:

$$\frac{1}{2} > \frac{\sqrt{\hat{\sigma}^2}}{D} + \sqrt{\frac{2\log(2/y)}{n-1}}$$

We plot the improvement our EBB offers in this domain, as shown in Figure 2. In this plot, a probability 0.5 bound is shown to shrink by approximately one third. More generally, we observe that our refinement of M&P's EBB is be uniformly tighter across a large range of values.

Third, a comparison is made of the further improvement in confidence over our EBB that can be achieved with perfect information about the variance; specifically, Bennett's inequality is used assuming $\hat{\sigma}^2 = \sigma^2$. This improvement is plotted in Figure 3, which shows that when the variance is small, uncertainty about the variance is the most detrimental to an EBB, such as ours. However, in general, going from our EBB to perfect variance information shrinks the bounds by about another third.

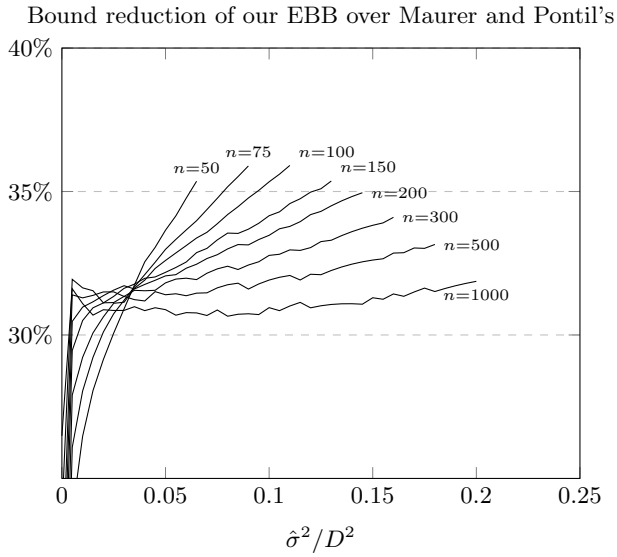

Fig. 2. The percent reduction of the 0.5 probability bound, that going from Maurer and Pontil's EBB to our EBB would achieve, for various n, in the domain valid for their EBB.

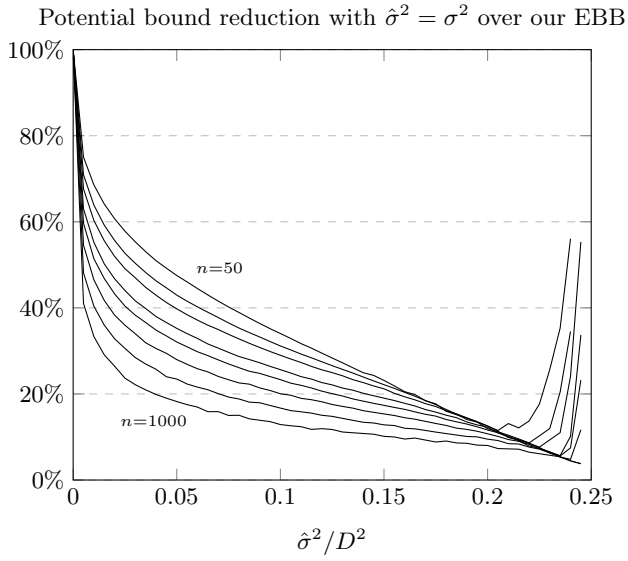

Fig. 3. The percent reduction in the 0.5 probability bound that going from our EBB to using Bennett's inequality (perfect variance information, $\hat{\sigma}^2 = \sigma^2$) achieves, for $n = 50, 75, 100, 150, 200, 300, 500, 1000$.

6 Application: Multi-Armed Bandits

One example use of concentration inequalities is in the context of the *upper-confidence bound* (UCB) method in *multi-armed bandit* (MAB) problems. In this section we consider the performance of UCB employing different concentration inequalities in an example MAB, in order to show the benefit of using our EBB.

6.1 MAB Problem Description

There are several variations of MAB problems, however the classic MAB [26] problem comprises a single bandit machine with K arms, each of which returns rewards that are independently drawn from an unknown distribution when it is pulled. In general, the MAB problem is to design an algorithm for sequentially choosing between the K arms in order to maximise the sum of the (initially unknown) stochastic rewards that each arm yields. Initially, a player must choose exploratory actions to learn about the rewards that each arm returns, before exploiting this information to choose the higher-valued arms. In this way, MABs illustrate finite horizon reinforcement learning dynamics, and is one of the clearest examples of the exploration-exploitation trade-off in machine learning.

Formally, at each time-step, n, a player has to choose which of the arms to pull. However, the player initially has no knowledge of the rewards of each arm, $k \in \mathcal{K}$, so it must learn these values in order to deduce a policy that maximises its sum of rewards. As argued above, in real-world applications, reward values are typically bounded, so we assume that each arm's reward distribution has bounded supports. Denote the mean of this distribution and the width of its support μ_k and D_k, respectively.

Let $A = \{a(1), a(2), \ldots\}$ be a finite sequence of arm pulls, where $a(n)$ is the arm pulled at time-step t, $a(n) \in \mathcal{K}$. Let $R(A)$ be the total return to the player from following the sequence A. The expectation of A is:

$$\mathbb{E}\left[R(A)\right] = \sum_{a(t) \in A} \mu_k$$

An optimal sequence of arm pulls, A^* is one that maximises the expression above, that is:

$$A^* = \arg\max_A \mathbb{E}\left[R(A)\right] = \arg\max_A \sum_{a(n) \in A} \mu_k$$

However, in order to determine A^*, we have to know the value of μ_k in advance, which we do not. Thus, A^* is a *theoretical* optimum value, which is not achievable in general. Instead, a typical approach to MABs is to define a *loss* or *regret* function, $L(A)$ for an arbitrary algorithm A:

$$L(A) = \mathbb{E}\left[R(A^*)\right] - \mathbb{E}\left[R(A)\right] \tag{8}$$

Using this regret function, the MAB problem is transformed to one of finding a sequence, A, that minimises $L(A)$.

6.2 Upper-Confidence Bound Methods

One well-known and effective strategy for the MAB is UCB [20]. Under UCB, at each iteration, n, the arm with the greatest upper confidence bound on the estimated mean of its reward as inferred from past rewards (at some confidence level) is selected. Specifically, the general form of UCB methods is to define a *confidence interval*, CI on the estimate of the mean:

$$\mathbb{P}\left(\mu - \hat{\mu} \geq CI\right) \leq y$$

where y is a confidence level, and then at each iteration, to select the arm with the greatest upper confidence bound given this confidence interval:

$$a(n) = \arg\max_{k \in \mathcal{K}} \left[\hat{\mu}_k(n) + CI_k(n)\right]$$

where $\hat{\mu}_k(n)$ and $CI_k(n)$ are the mean estimate and confidence interval at time-step n, respectively. In this way, the initial selection of arms is driven by the degree of uncertainty about their rewards, as captured by using the confidence bounds, while over time, the best performing arms are selected more often.

UCB methods can be categorised by the specific type of bandit problem they apply to, and also the method used to infer the confidence interval. One typical UCB method uses Hoeffding's inequality to set the confidence interval [6], where Hoeffding's inequality is given by:

$$\mathbb{P}\left(\mu - \hat{\mu} \geq \sqrt{\frac{D^2 \log(1/y)}{2n}}\right) \leq y. \tag{9}$$

Additionally we consider the EBB type UCB method developed by [4] per their inequality:

$$\mathbb{P}\left(\mu - \hat{\mu} \geq \sqrt{\frac{\hat{\sigma}^2 \log(3/t)}{2n}} + \frac{3D \log(3/t)}{2n}\right) \leq t. \tag{10}$$

We also consider a UCB method with the EBB developed by M&P [22], particularly utilizing the bound of inequality (7) (in Section 5). All three are compared to UCB employing our EBB to define the upper confidence bound. Additionally, we also compare to randomly choosing actions, for a naïve baseline. In all cases we selected UCB to minimise a probability 0.5 bound.

We used a confidence level of 0.5 in all cases simply as a representative of a mid-range bound, but note that potentially some different dynamics could occur with the selection of more extreme bounds (i.e. close to 0 or 1).

For the application of our EBB we hand-tuned a function approximating our EBB's numerical probability 0.5 bound:

$$\mathbb{P}\left(\mu - \hat{\mu} \geq \frac{D}{\sqrt{n}} \min\left[\sqrt{2\log 2}, \left(\begin{array}{c}\frac{3}{5}\sqrt{\min\left[1, \frac{\hat{\sigma}^2}{D^2} + \frac{25}{n}\right]} \\ + \ln\left(\max\left[1, n\left(1 - \frac{\hat{\sigma}^2}{D^2}\right)\right]\right)\end{array}\right)^{-4}\right]\right) \lessapprox 0.5 \tag{11}$$

The process of creating the above expression involved plotting the numerical data, and manually fitting an approximate symbolic expression. This expression was used *in situ* to simplify the application of our EBB in the bandit context. However the numeric data itself may have been calculated and used directly, at the cost of longer compute times.

6.3 Problem instances and results

In the example bandit problems considered here, the number of arms is $K = 8$, each of which yield rewards of between 0 and 1. For each arm k, there is a unique α_k and β_k parameters of its beta distribution over rewards, and for each realization of the problem, these α_k and β_k are drawn uniformly from between 0 and 3. For these problems, we used the different confidence bound approaches, and measured their performance in terms of regret, defined in (8). As noted above, regret is a measure of the performance of bandit algorithms identified by the expected loss of selecting an arm against choosing only the ideal arm. The regret of the different methods of choosing actions was estimated as the average regret obtained across 100,000 instances of this bandit problem. We computed the average regret of these methods over finite arm-pulling budgets, N, in order to assess the algorithm's finite-time performance.

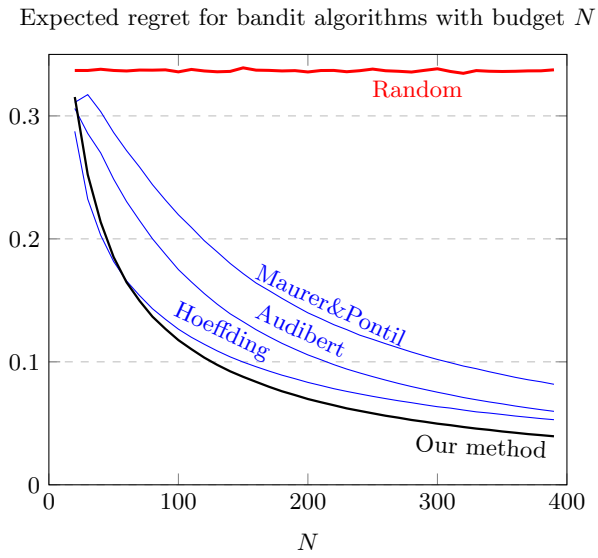

Fig. 4. The expected regret of bandit algorithms and a baseline method in the example bandit problem: UCB method using our bound (11), Hoeffding's, Audibert et.al's, and Maurer & Pontil's inequalities; and method of uniform randomly choosing an arm.

The the performance of the four methods and the naïve baseline are shown in Figure 4. From this figure, we see that minimizing an upper confidence bound utilizing our inequality (11) results in best performance (lowest regret), except marginally in the region of very small sample budgets.

It is somewhat surprising to see the more complicated EBB methods [22, 4] perform worse than the much simpler Hoeffding's inequality. As EBB inequalities are specifically constructed in a way to incorporate the estimate of the variance, then the potential advantage comes when there are sufficiently many samples for reliable variance estimation. However a bound without this construction (such as Hoeffding's inequality) may be tighter and more effective for small/medium sample budgets. As expected, the random method of choosing arms had a constant expected average regret across action budgets, as it does not learn with additional samples of the arms' rewards.

7 Conclusion

In this paper, we have extended existing work on concentration inequalities to derive a new and stronger EBB. Our EBB has many applications, in any setting where a mean value must be estimated with confidence, such as bandit problems. Our EBB was shown to tighten known EBB-based confidence intervals by about a third, thereby improving the value of these types of concentration inequalities. This value was demonstrated in a MAB problem, where using our EBB in a UCB algorithm was shown to improve online learning performance.

A Proofs

A.1 Small Proofs

Proof (Proof of Probability Union - Lemma 1). For any events A and B
$\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$, hence for events $a > b$ and $b > c$:
$\mathbb{P}((a > b) \cup (b > c)) \leq \mathbb{P}(a > b) + \mathbb{P}(b > c)$
If $a > c$, then $(a > b) \cup (b > c)$ is true irrespective of b, so:
$\mathbb{P}(a > c) \leq \mathbb{P}((a > b) \cup (b > c))$ ■

Proof (Proof of Variance Decomposition - Lemma 2). By expanding and $\hat{\sigma}^2$:
$\hat{\sigma}^2 = \frac{1}{n-1} \sum_i \left(x_i - \frac{1}{n} \sum_j x_j \right)^2 = \frac{1}{n-1} \left(\sum_i x_i^2 - \frac{1}{n} \sum_{i,j} x_i x_j \right) = \frac{n}{n-1} \left(\hat{\sigma}_0^2 - \hat{\mu}^2 \right)$ ■

Proof (Proof of Chernoff Bound - Lemma 3).
$\mathbb{P}(\hat{\mu} \geq t) = \mathbb{P}\left(\exp\left(s \sum_{i=1}^n x_i \right) \geq \exp(snt) \right)$
$\leq \mathbb{E}\left[\exp\left(s \sum_{i=1}^n x_i \right) \right] \exp(-snt) \leq \mathbb{E}\left[\exp\left(sX \right) \right]^n \exp(-snt)$
using Markov's inequality and the i.i.d of the samples, respectively. ■

A.2 A Proof of Bennett's inequality

Theorem 8 (Parabola Fitting). *For $b > 0$, $a < b$ and $z > 0$, there exists an α, β, γ such that: $\alpha x^2 + \beta x + \gamma \geq \exp(x)$ for all $a \leq x \leq b$, and:*
$z\alpha + \gamma = (z\exp(b) + b^2\exp(-z/b))(z + b^2)^{-1}$.

Proof. A example parabola $\alpha x^2 + \beta x + \gamma$ which that satisfies these requirements tangentially touches the exponential curve at one point (at $x = f < b$) and intersects it at another (at $x = b$), as illustrated in Figure 5. Thus the parabola's intersection at $x = b$ and its tangential intersection at $x = f$ can be written in matrix algebra:

$$\begin{bmatrix}\alpha\\ \beta\\ \gamma\end{bmatrix} = \begin{bmatrix}b^2 & b & 1\\ f^2 & f & 1\\ 2f & 1 & 0\end{bmatrix}^{-1}\begin{bmatrix}\exp(b)\\ \exp(f)\\ \exp(f)\end{bmatrix}$$

This gives our parabola parameters α, β, γ, in terms of f and b, hence:
$z\alpha + \gamma = (((z + fb - b)(f - b - 1) - b)e^f + (f^2 + z)e^b)(b - f)^{-2}$
Minimizing with respect to f occurs at $f = \frac{-z}{b}$ and gives the result. ■

Proof (Proof of Bennett's inequlity – Lemma 4). As random variable X is bounded $a \leq X \leq b$, for any $s > 0$, by Theorem 8, there exist parameters α, β, γ such that, $\alpha s^2 X^2 + \beta s X + \gamma \geq \exp(sX)$ is always satisfied, hence for these we have:
$$\begin{aligned}\mathbb{E}\left[\exp(sX)\right] &\leq \mathbb{E}[\alpha s^2 X^2 + \beta s X + \gamma] \leq \alpha s^2\,\mathbb{E}[X^2] + \gamma \leq \alpha s^2\sigma^2 + \gamma\\ &\leq (\sigma^2\exp(sb) + b^2\exp(-s\sigma^2/b))(\sigma^2 + b^2)^{-1}\end{aligned}$$
Hence by application of lemma 3:
$\mathbb{P}(\hat{\mu} \geq t) \;\leq\; (\sigma^2\exp(sb) + b^2\exp(-s\sigma^2/b))^n((\sigma^2 + b^2)\exp(st))^{-n}$
and finding the minimum with respect to s completes the proof. ■

A.3 Remaining Proofs

Proof (Proof of Sample Square Bound - Lemma 5). There exist parameters α, γ such for all $a \leq X \leq b$ that $\alpha X^2 + \gamma \geq \exp(-qX^2)$ whence:
$\mathbb{E}[\exp(-qX^2)] \leq \mathbb{E}[\alpha x^2 + \gamma] \leq \alpha\sigma^2 + \gamma$
With $d = \max(b, -a)$, we choose (Fig 6) $\alpha = (\exp(-qd^2) - 1)d^{-2}$ and $\gamma = 1$
Then applying lemma 3 to the mean of the negated sample squares gives:
$\mathbb{P}(-\hat{\sigma}_0^2 \geq t) \leq \left(\frac{\sigma^2}{d^2}\exp(-qd^2) + 1 - \frac{\sigma^2}{d^2}\right)^n \exp(-qnt)$
Substituting t for $y - \sigma^2$ and minimizing with q completes the proof. ■

Proof (Proof of Sample Variance Bound - Lemma 6). By Lemmas 5 and 2:
$\mathbb{P}\left(\sigma^2 - \hat{\sigma}^2 > \frac{n}{n-1}\left(\hat{\mu}^2 + y - \frac{1}{n}\sigma^2\right)\right) \leq H_2^n\left(\frac{\sigma^2}{d^2}, \frac{y}{d^2}\right)$
Also, by manipulating the inner inequality of equation 2:
$\mathbb{P}\left(\frac{n}{n-1}\left(\hat{\mu}^2 + y - \frac{1}{n}\sigma^2\right) \geq \frac{n}{n-1}\left(r^2 + y - \frac{1}{n}\sigma^2\right)\right) \leq H_1^n\left(\frac{\sigma^2}{b^2}, \frac{r}{b}\right) + H_1^n\left(\frac{\sigma^2}{a^2}, \frac{-r}{a}\right)$
Applying lemma 1 to the above two equations gives:
$\mathbb{P}\left(\sigma^2 - \hat{\sigma}^2 > \frac{n}{n-1}\left(r^2 + y - \frac{1}{n}\sigma^2\right)\right) \leq H_2^n\left(\frac{\sigma^2}{d^2}, \frac{y}{d^2}\right) + H_1^n\left(\frac{\sigma^2}{b^2}, \frac{r}{b}\right) + H_1^n\left(\frac{\sigma^2}{a^2}, \frac{-r}{a}\right)$
For $w = \frac{n}{n-1}\left(r^2 + y - \frac{1}{n}\sigma^2\right)$ there is a range of possible $r, y > 0$ which we parameterise by value ϕ, such that $0 \leq \phi \leq 1$:

$y(\phi) = (1-\phi)\left(\frac{n-1}{n}w + \frac{1}{n}\sigma^2\right)$ and $r(\phi)^2 = \phi\left(\frac{n-1}{n}w + \frac{1}{n}\sigma^2\right)$

Thus:

$\mathbb{P}\left(\sigma^2 - \hat{\sigma}^2 > w\right) \leq H_2^n\left(\frac{\sigma^2}{d^2}, \frac{y(\phi)}{d^2}\right) + H_1^n\left(\frac{\sigma^2}{b^2}, \frac{r(\phi)}{b}\right) + H_1^n\left(\frac{\sigma^2}{a^2}, \frac{-r(\phi)}{a}\right)$

The result of this proof follows by taking the minimum over ϕ. ■

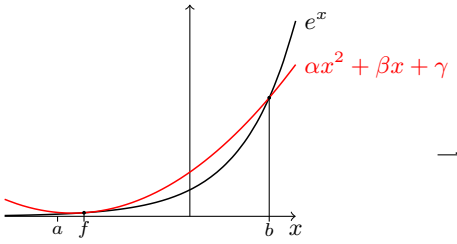

Fig. 5. A parabola parametarised by touching and intercepting points f, b above an exponential curve for all $a \leq x \leq b$

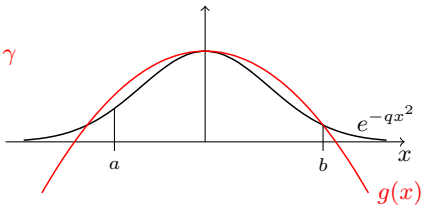

Fig. 6. $g(x) = (e^{-qd^2} - 1)d^{-2}x^2 + 1$ over function $f(x) = e^{-qx^2}$ for all $a \leq x \leq b$ where $d = \max(b, -a)$; in the case $a = -1, b = 1.3, q = 1$

References

1. Agarwal, A., Dekel, O., Xiao, L.: Optimal algorithms for online convex optimization with multi-point bandit feedback. In: 23rd Annual Conf. Learning Theory (COLT'10) (2010)
2. Aggarwal, C.C.: Data Mining: The Textbook, chap. Outlier analysis, pp. 237–263. Springer Publishing Company, Incorporated (2015)
3. Audibert, J.Y., Bubeck, S.: Minimax policies for adversarial and stochastic bandits. In: 22nd Annual Conf. Learning Theory (COLT'09) (2009)
4. Audibert, J.Y., Munos, R., Szepesvári, C.: Tuning bandit algorithms in stochastic environments. In: Hutter, M., Servedio, R.A., Takimoto, E. (eds.) Algorithmic Learning Theory. pp. 150–165. Springer Berlin Heidelberg, Berlin, Heidelberg (2007)
5. Auer, P., Cesa-Bianchi, N., Freund, Y., Schapire, R.: The non-stochasticmulti-armed bandit problem. SIAM Journal on Computing **31**(1), 48–77 (2003)
6. Auer, P., Cesa-Bianchi, N., Fischer, P.: Finite-time analysis of the multiarmed bandit problem. Machine Learning **47**(2), 235–256 (May 2002)
7. Bennett, G.: Probability inequalities for the sum of independent random variables. Journal of the American Statistical Association **57**(297), 33–45 (1962)
8. Bentkus, V., Juškevičius, T.: Bounds for tail probabilities of martingales using skewness and kurtosis. Lithuanian Mathematical Journal **48**(1), 30–37 (Jan 2008)
9. Bercu, B., Delyon, B., Rio, E.: Concentration inequalities for sums and martingales. Springer Briefs in Mathematics, Springer (2015)
10. Bernstein, S.N.: On a modification of Chebyshev's inequality and of the error formula of Laplace. Uchenye Zapiski Nauch.-Issled. Kaf. Ukraine, Sect. Math **1**, 38–48 (1924)

11. Bhaduri, M., Zhan, J., Chiu, C., Zhan, F.: A novel online and non-parametric approach for drift detection in big data. IEEE Access **5**, 15883–15892 (2017)
12. Bienaymé, I.J.: Considérations àl'appui de la découverte de Laplace. Comptes Rendus de l'Académie des Sciences **37**, 309–324 (1853)
13. Boucheron, S., Lugosi, G., Bousquet, O.: Concentration inequalities. In: Bousquet, O., von Luxburg, U., Rätsch, G. (eds.) Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2 - 14, 2003, Tübingen, Germany, August 4 - 16, 2003, Revised Lectures, pp. 208–240. Springer Berlin Heidelberg, Berlin, Heidelberg (2004)
14. Boucheron, S., Lugosi, G., Massart, P.: Concentration inequalities using the entropy method. The Annals of Probability **31**(3), 1583–1614 (2003)
15. Cesa-Bianchi, N., Lugosi, G.: Prediction, Learning, and Games. Cambridge University Press (2006)
16. Efron, B., Stein, C.: The jackknife estimate of variance. Annals of Statistics **9**(3), 586–596 (05 1981)
17. Flaxman, A., Kalai, A., McMahan, B.: Online convex optimization in the bandit setting: Gradient descent without a gradient. In: Proc. 16th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'05). p. 385–394 (2005)
18. Hoeffding, W.: Probability inequalities for sums of bounded random variables. Journal of the American Statistical Association **58**(301), 13–30 (Mar 1963)
19. Kifer, D., Ben-David, S., Gehrke, J.: Detecting change in data streams. In: Proc. 30th Int. Conf. Very Large Data Bases (VLDB'04). pp. 180–191. VLDB '04 (2004)
20. Lai, T., Robbins, H.: Asymptotically efficient adaptive allocation rules. Advances in Applied Mathematics **6**(1), 4–22 (Mar 1985)
21. Maurer, A.: Concentration inequalities for functions of independent variables. Random Structures & Algorithms **29**(2), 121–138 (2006)
22. Maurer, A., Pontil, M.: Empirical Bernstein bounds and sample variance penalization. stat. Proceedings of the 22nd Annual Conf. Learning Theory (COLT'09) (June 2009)
23. Mnih, V., Szepesvári, C., Audibert, J.Y.: Empirical Bernstein stopping. In: Proc. 25th Int. Conf. Machine Learning (ICML'08). pp. 672–679 (2008)
24. Pinelis, I.: On the Bennett-Hoeffding inequality. Annales de i'Institut Henri Poincaré - Probabilités et Statistiques **50**(1), 15–27 (2014)
25. Rehman, M.Z., Li, T., Li, T.: Exploiting empirical variance for data stream classification. Journal of Shanghai Jiaotong University (Science) **17**(2), 245–250 (Apr 2012)
26. Robbins, H.: Some aspects of the sequential design of experiments. Bulletin of the AMS **55**, 527–535 (1952)
27. Sharma, R., Gupta, M., Kapoor, G.: Some better bounds on the variance with applications. Journal of Mathematical Inequalities **4**(3), 355–363 (2010)
28. Talagrand, M.: Concentration of measure and isoperimetric inequalities in product spaces. Publications Mathématiques de l'Institut des Hautes Études Scientifiques **81**(1), 73–205 (Dec 1995)
29. Talagrand, M.: The missing factor in Hoeffding's inequalities. Annales de l'Institut Henri Poincare, Probability and Statistics **31**(4), 689–702 (1995)
30. Thomas, P.S., Theocharous, G., Ghavamzadeh, M.: High-confidence off-policy evaluation. In: Proc. 29th AAAI Conf. Artificial Intelligence (AAAI'15). pp. 3000–3006 (2015)
31. Tran-Thanh, L., Chapman, A.C., Rogers, A., Jennings, N.R.: Knapsack based optimal policies for budget-limited multi-armed bandits. In: Proc. 26th AAAI Conf. Artificial Intelligence (AAAI'12). pp. 1134–1140. AAAI'12 (2012)