Optimal Non-Asymptotic Lower Bound on the Minimax Regret of Learning with Expert Advice

Francesco Orabona, David Pal
francesco@orabona.com, dpal@yahoo-inc.com
Yahoo Labs, New York, NY, USA
November 9, 2015

Abstract

We prove non-asymptotic lower bounds on the expectation of the maximum of $d$ independent Gaussian variables and the expectation of the maximum of $d$ independent symmetric random walks. Both lower bounds recover the optimal leading constant in the limit. A simple application of the lower bound for random walks is an (asymptotically optimal) non-asymptotic lower bound on the minimax regret of online learning with expert advice.

1 Introduction

Let $X_1, X_2, \ldots, X_d$ be i.i.d. Gaussian random variables $\mathcal{N}(0, \sigma^2)$. It is easy to prove that (see Appendix A)

$$\mathbb{E} \left[ \max_{1 \leq i \leq d} X_i \right] \leq \sigma \sqrt{2 \ln d} \quad \text{for any } d \geq 1. \tag{1}$$

It is also well known that

$$\lim_{d \to \infty} \mathbb{E} \left[ \max_{1 \leq i \leq d} X_i \right] = 1. \tag{2}$$

In section 2, we prove a non-asymptotic $\Omega(\sigma \sqrt{\log d})$ lower bound on $\mathbb{E} \left[ \max_{1 \leq i \leq d} X_i \right]$.

Discrete analog of a Gaussian random variable is the symmetric random walk. Recall that a random walk $Z^{(n)}$ of length $n$ is a sum $Z^{(n)} = Y_1 + Y_2 + \cdots + Y_n$ of $n$ i.i.d. Rademacher variables, which have probability distribution $\Pr[Y_i = +1] = \Pr[Y_i = -1] = 1/2$. We consider $d$ independent symmetric random walks $Z^{(n)}_1, Z^{(n)}_2, \ldots, Z^{(n)}_d$ of length $n$. Analogously to (1), it is easy to prove that (see Appendix A)

$$\mathbb{E} \left[ \max_{1 \leq i \leq d} Z^{(n)}_i \right] \leq \sqrt{2n \ln d} \quad \text{for any } n \geq 0 \text{ and any } d \geq 1. \tag{3}$$

Note that $\sigma^2$ in (1) is replaced by $\text{Var}(Z^{(n)}_i) = n$. By central limit theorem $\frac{Z^{(n)}_i}{\sqrt{n}}$ as $n \to \infty$ converges in distribution to $\mathcal{N}(0, 1)$. From this fact, it possible to prove the analog of (2).

$$\lim_{d \to \infty} \lim_{n \to \infty} \mathbb{E} \left[ \max_{1 \leq i \leq d} Z^{(n)}_i \right] = 1. \tag{4}$$
We prove a non-asymptotic $\Omega(\sqrt{n\log d})$ lower bound on $\mathbb{E}\left[ \max_{1 \leq i \leq d} Z_i^{(n)} \right]$. Same as for the Gaussian case, the leading term of the lower bound is asymptotically $\sqrt{2n \ln d}$ matching (4).

In section 4 we show a simple application of the lower bound on $\mathbb{E}\left[ \max_{1 \leq i \leq d} Z_i^{(n)} \right]$ to the problem of learning with expert advice. This problem was extensively studied in the online learning literature; see [Cesa-Bianchi and Lugosi, 2006]. Our bound is optimal in the sense that for large $d$ and large $n$ it recovers the right leading constant.

2 Maximum of Gaussians

Crucial step towards lower bounding $\mathbb{E}\left[ \max_{1 \leq i \leq d} X_i \right]$ is a good lower bound on the tail $\Pr[X_i \geq x]$ of a single Gaussian. The standard way of deriving such bounds is via bounds on the so-called Mill’s ratio. Mill’s ratio of a random variable $X$ with density function $f(x)$ is the ratio $\frac{\Pr[X \geq x]}{f(x)}$. It clear that a lower bound on the Mill’s ratio yields a lower bound on the tail $\Pr[X > x]$.

Without loss of generality it suffices to lower bound the Mill’s ratio of $N(0, 1)$, since Mill’s ratio of $N(0, \sigma^2)$ can be obtained by rescaling. Recall that probability density of $N(0, 1)$ is $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ and its cumulative distribution function is $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$. The Mill’s ratio for $N(0, 1)$ can be expressed as $\frac{1 - \Phi(x)}{\phi(x)}$. A lower bound on Mill’s ratio of $N(0, 1)$ was proved by Boyd [1959].

**Lemma 1** (Mill’s ratio for standard Gaussian [Boyd 1959]). For any $x \geq 0$,

\[
1 - \Phi(x) = \exp \left( \frac{x^2}{2} \right) \int_{x}^{\infty} \exp \left( -\frac{t^2}{2} \right) dt \geq \frac{\pi}{(\pi - 1)x + \sqrt{x^2 + 2\pi}} \geq \frac{\pi}{\pi x + \sqrt{2\pi}}.
\]

The second inequality in Lemma 1 is our simplification of Boyd’s bound. It follows by setting $a = \sqrt{2\pi}$ and $b = x$. By a simple algebra it is equivalent to the inequality $a + b \leq \sqrt{a^2 + b^2}$ which holds for any $a, b \geq 0$.

**Corollary 2** (Lower Bound on Gaussian Tail). Let $X \sim N(0, \sigma^2)$ and $x \geq 0$. Then,

\[
\Pr[X \geq x] \geq \exp \left( -\frac{x^2}{2\sigma^2} \right) \frac{1}{\sqrt{2\pi \sigma^2} + 2}.
\]

**Proof.** We have

\[
\Pr[X \geq x] = \frac{1}{\sigma\sqrt{2\pi}} \int_{x}^{\infty} \exp \left( -\frac{t^2}{2\sigma^2} \right) dt \\
= \frac{1}{\sqrt{2\pi}} \int_{\frac{x}{\sigma}}^{\infty} \exp \left( -\frac{t^2}{2} \right) dt \\
\geq \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{x^2}{2\sigma^2} \right) \frac{\pi}{\pi x + \sqrt{2\pi}} \quad \text{(by Lemma 1)}.
\]

Equipped with the lower bound on the tail, we prove a lower bound on the maximum of Gaussians.

**Theorem 3** (Lower Bound on Maximum of Independent Gaussians). Let $X_1, X_2, \ldots, X_d$ be independent Gaussian random variables $N(0, \sigma^2)$. For any $d \geq 2$,

\[
\mathbb{E}\left[ \max_{1 \leq i \leq d} X_i \right] \geq \sigma \left( 1 - \exp \left( -\frac{\sqrt{\ln d}}{6.35} \right) \left( \sqrt{2\ln d - 2 \ln \ln d + \frac{2}{\pi}} - \sqrt{\frac{2}{\pi}} \right) - \sqrt{\frac{2}{\pi}} \sigma \right) \geq 0.13\sigma\sqrt{\ln d - 0.7\sigma}.
\]

\footnote{Mill’s ratio has applications in economics. A simple is problem where Mill’s ratio shows up is the problem of setting optimal price for a product. Given a distribution prices that customers are willing to pay, the goal is to choose the price that brings the most revenue.}
Proof. Let $A$ be the event that at least one of the $X_i$ is greater than $C\sigma\sqrt{\ln d}$ where $C = C(d) = \sqrt{2 - \frac{2\ln \ln d}{\ln d}}$. We denote by $\overline{A}$ the complement of this event. We have

$$
\mathbb{E} \left[ \max_{1 \leq i \leq d} X_i \right] = \mathbb{E} \left[ \max_{1 \leq i \leq d} X_i \mid A \right] \cdot \Pr[A] + \mathbb{E} \left[ \max_{1 \leq i \leq d} X_i \mid \overline{A} \right] \cdot \Pr[\overline{A}]
$$

$$
\geq \mathbb{E} \left[ \max_{1 \leq i \leq d} X_i \mid A \right] \cdot \Pr[A] + \mathbb{E} \left[ X_1 \mid \overline{A} \right] \cdot \Pr[\overline{A}]
$$

$$
= \mathbb{E} \left[ \max_{1 \leq i \leq d} X_i \mid A \right] \cdot \Pr[A] + \mathbb{E} \left[ X_1 \mid X_1 \leq C\sigma\sqrt{\ln d} \right] \cdot \Pr[\overline{A}]
$$

$$
\geq \mathbb{E} \left[ \max_{1 \leq i \leq d} X_i \mid A \right] \cdot \Pr[A] + \mathbb{E}[X_1 \mid X_1 \leq 0] \cdot \Pr[\overline{A}]
$$

$$
\geq \mathbb{E} \left[ \max_{1 \leq i \leq d} X_i \mid A \right] \cdot \Pr[A] - \sigma \sqrt{\frac{2}{\pi}} \cdot \Pr[\overline{A}]
$$

$$
\geq C\sigma\sqrt{\ln d} \cdot \Pr[A] - \sigma \sqrt{\frac{2}{\pi}} (1 - \Pr[A])
$$

$$
= \sigma \left( C\sqrt{\ln d} + \sqrt{\frac{2}{\pi}} \right) \Pr[A] - \sigma \sqrt{\frac{2}{\pi}} 
$$

where we used that $\mathbb{E}[X_1 \mid X_1 \leq 0] = \frac{1}{\Pr[X_1 \leq 0]} \int_{-\infty}^{0} \frac{x}{\sigma \sqrt{2\pi}} \exp \left( -\frac{x^2}{2\sigma^2} \right) = -\sigma \sqrt{\frac{2}{\pi}}$.

It remains to lower bound $\Pr[A]$, which we do as follows

$$
\Pr[A] = 1 - \Pr[\overline{A}]
$$

$$
= 1 - \left( \Pr \left[ X_1 \leq C\sigma\sqrt{\ln d} \right] \right)^d
$$

$$
= 1 - \left( 1 - \Pr \left[ X_1 > C\sigma\sqrt{\ln d} \right] \right)^d
$$

$$
\geq 1 - \exp \left( -d \cdot \Pr \left[ X_1 \geq C\sigma\sqrt{\ln d} \right] \right)
$$

$$
\geq 1 - \exp \left( -d \exp \left( -\frac{C^2 \ln d}{2} \right) \frac{1}{\sqrt{2\pi} C \sqrt{\ln d} + 2} \right)
$$

$$
= 1 - \exp \left( -\frac{d^1 - \frac{C^2 \ln d}{2}}{C \sqrt{2\pi} \ln d + 2} \right). 
$$

where in the first inequality we used the elementary inequality $1 - x \leq \exp(-x)$ valid for all $x \in \mathbb{R}$.

Since $C = \sqrt{2 - \frac{2\ln \ln d}{\ln d}}$ we have $d^1 - \frac{C^2 \ln d}{2} = \ln d$. Substituting this into (8), we get

$$
\Pr[A] \geq 1 - \exp \left( -\frac{\ln d}{C \sqrt{2\pi} \ln d + 2} \right) = 1 - \exp \left( -\frac{\sqrt{\ln d}}{C \sqrt{2\pi} + 2} \right).
$$

The function $C(d)$ is decreasing on the interval $[1, e^c]$, increasing on $[e^c, \infty)$, and $\lim_{d \to \infty} C(d) = \sqrt{2}$. From these properties we can deduce that $C(d) \leq \max \{ C(2), \sqrt{2} \} \leq 1.75$ for any $d \in [2, \infty)$. Therefore, $C \sqrt{2\pi} + 2 \leq 6.35$ and hence

$$
\Pr[A] \geq 1 - \exp \left( -\frac{\sqrt{\ln d}}{6.35} \right).
$$
Inequalities (7) and (10) together imply bound (5). Bound (6) is obtained from (5) by noticing that

$$\sigma \left( 1 - \exp \left( -\frac{\sqrt{\ln d}}{6.35} \right) \right) \left( \sqrt{2 \ln d - 2 \ln \ln d + \frac{2}{\pi}} \right) - \sqrt{\frac{2}{\pi} \sigma}$$

$$= \sigma \left( 1 - \exp \left( -\frac{\sqrt{\ln d}}{6.35} \right) \right) \sqrt{2 \ln d - 2 \ln \ln d - \exp \left( -\frac{\ln d}{6.35} \right) \sqrt{\frac{2}{\pi} \sigma}}$$

$$\geq 0.1227 \cdot \sigma \sqrt{2 \ln d - 2 \ln \ln d - 0.7 \sigma}$$

$$= 0.1227 \cdot \sigma \sqrt{\ln d} \cdot C(d) - 0.7 \sigma$$

where we used that $\exp \left( -\frac{\sqrt{\ln d}}{6.35} \right) \leq 0.8773$ for any $d \geq 2$. Since $C(d)$ has minimum at $d = e^\sigma$, it follows that $C(d) \geq C(e^\sigma) = \sqrt{2 - \frac{2}{\pi}} \geq 1.1243$ for any $d \geq 2$. $\square$

### 3 Maximum of Random Walks

The general strategy for proving a lower bound on $\mathbf{E} \left[ \max_{1 \leq i \leq d} Z_i^{(n)} \right]$ is the same as in the previous section. The main task is to lower bound the tail $\Pr[Z^{(n)} \geq x]$ of a symmetric random walk $Z^{(n)}$ of length $n$. Note that

$$B_n = \frac{Z^{(n)} + n}{2}$$

is a Binomial random variable $B(n, \frac{1}{2})$. We follow the same approach used in Orabona [2013]. First we lower bound the tail $\Pr[B_n \geq k]$ with McKay [1989, Theorem 2].

**Lemma 4 (Bound on Binomial Tail).** Let $n, k$ be integers satisfying $n \geq 1$ and $\frac{n}{2} \leq k \leq n$. Define $x = \frac{2k-n}{\sqrt{n}}$. Then, $B_n \sim B(n, \frac{1}{2})$ satisfies

$$\Pr[B_n \geq k] \geq \sqrt{n} \binom{n-1}{k-1} 2^{-n} \frac{1 - \Phi(x)}{\phi(x)} .$$

We lower bound the binomial coefficient $\binom{n-1}{k-1}$ using Stirling’s approximation of the factorial. The lower bound on the binomial coefficient will be expressed in terms of Kullback-Leibler divergence between two Bernoulli distributions, Bernoulli($p$) and Bernoulli($q$). Abusing notation somewhat, we write the divergence as

$$D(p||q) = p \ln \left( \frac{p}{q} \right) + (1 - p) \ln \left( \frac{1 - p}{1 - q} \right) .$$

The result is the following lower bound on the tail of Binomial.

**Theorem 5 (Bound on Binomial Tail).** Let $n, k$ be integers satisfying $n \geq 1$ and $\frac{n}{2} \leq k \leq n$. Define $x = \frac{2k-n}{\sqrt{n}}$. Then, $B_n \sim B(n, \frac{1}{2})$ satisfies

$$\Pr[B_n \geq k] \geq \exp \left( -nD \left( \frac{k}{n} \middle|\middle| \frac{1}{2} \right) \right) \frac{1 - \Phi(x)}{\phi(x)} .$$

**Proof.** Lemma 4 implies that

$$\Pr[B_n \geq k] \geq \sqrt{n} \binom{n-1}{k-1} 2^{-n} \frac{1 - \Phi(x)}{\phi(x)} .$$

Since $k \geq 1$, we can write the binomial coefficient as

$$\binom{n-1}{k-1} = \frac{k}{n} \binom{n}{k}$$
We bound the binomial coefficient \( \binom{n}{k} \) by using Stirling’s formula for the factorial. We use explicit upper and lower bounds due to [Robbins 1955] valid for any \( n \geq 1 \),

\[
\sqrt{2\pi n} \left( \frac{n}{e} \right)^n < n! < \exp \left( \frac{1}{12} \right) \sqrt{2\pi n} \left( \frac{n}{e} \right)^n.
\]

Using the Stirling’s approximation, for any \( 1 \leq k \leq n - 1 \),

\[
\binom{n}{k} = \frac{n!}{k!(n-k)!} > \frac{\sqrt{2\pi n} \ n^n e^{-n}}{\sqrt{2\pi k \ k^k e^{-k} e^{1/12} \cdot \sqrt{2\pi (n-k) (n-k)^{n-k} e^{-(n-k) e^{1/12}}}}}
\]

\[
= \frac{1}{\exp \left( \frac{1}{12} \right) \sqrt{2\pi}} \left( \frac{n}{n-k} \right)^{n-k} \left( \frac{n}{k} \right)^k \sqrt{\frac{n}{k(n-k)}}
\]

\[
= \frac{1}{\exp \left( \frac{1}{12} \right) \sqrt{2\pi}} 2^n \exp \left( -n \cdot D \left( \frac{k}{n} \parallel \frac{1}{2} \right) \right) \frac{1 - \Phi(x)}{\phi(x)}
\]

where in the equality we used the definition of \( D(p||q) \). Combining all the inequalities, gives

\[
\Pr \left[ B_n \geq 2k - n \right] \geq \sqrt{\frac{1}{n}} \frac{1}{\exp \left( \frac{1}{12} \right) \sqrt{2\pi}} 2^n \exp \left( -n \cdot D \left( \frac{k}{n} \parallel \frac{1}{2} \right) \right) \frac{1 - \Phi(x)}{\phi(x)}
\]

\[
= \frac{1}{\exp \left( \frac{1}{12} \right) \sqrt{2\pi}} \exp \left( -n \cdot D \left( \frac{k}{n} \parallel \frac{1}{2} \right) \right) \frac{1 - \Phi(x)}{\phi(x)}
\]

\[
= \frac{1}{\exp \left( \frac{1}{12} \right) \sqrt{2\pi}} \exp \left( -n \cdot D \left( \frac{k}{n} \parallel \frac{1}{2} \right) \right) \frac{1 - \Phi(x)}{\phi(x)}
\]

for \( \frac{n}{2} \leq k \leq n - 1 \). For \( k = n \), we verify the statement of the theorem by direct substitution. The left hand side is \( \Pr[B^{(n)} \geq n] = 2^{-n} \). Since \( e^{-n D(1||\frac{1}{2})} = 2^{-n} \) and \( x = \sqrt{n} \geq 1 \), it’s easy to see that the right hand side is smaller than \( 2^{-n} \).

For \( k = n/2 + xn \), the divergence \( D \left( \frac{k}{n} \parallel \frac{1}{2} \right) = D \left( \frac{x}{n} \parallel \frac{1}{2} \right) \) can be approximated by \( 2x^2 \). We define the function \( \psi : [-\frac{1}{2}, \frac{1}{2}] \rightarrow \mathbb{R} \) as

\[
\psi(x) = \frac{D \left( \frac{x}{n} \parallel \frac{1}{2} \right)}{2x^2}.
\]

It is the ratio of the divergence and the approximation. The function \( \psi(x) \) satisfies the following properties:

- \( \psi(x) = \psi(-x) \)
- \( \psi(x) \) is decreasing on \( [-\frac{1}{2}, 0] \) and increasing on \( [0, \frac{1}{2}] \)
- minimum value is \( \psi(0) = 1 \)
- maximum value is \( \psi(\frac{1}{2}) = \psi(-\frac{1}{2}) = 2 \ln(2) \leq 1.3863 \)

Using the definition of \( \psi(x) \) and Theorem 5 we have the following Corollary.

**Corollary 6.** Let \( n \geq 1 \) be a positive integer and let \( t \in [1, \frac{n}{2} + 1] \) be a real number. Then \( B_n \sim B(n, \frac{1}{2}) \) satisfies

\[
\Pr \left[ B_n \geq \frac{1}{2}n + t - 1 \right] \geq \exp \left( -\frac{1}{6} \right) \exp \left( -2\psi \left( \frac{t}{n} \right) \frac{t^2}{n} \right) \frac{1}{\sqrt{2\pi \frac{1}{\psi^2(n) + 2}}}.
\]
Proof. By Theorem 5 and Lemma 1 we have

\[
\Pr \left[ Z \geq \frac{1}{2} n + t - 1 \right] = \Pr \left[ Z \geq \left( \frac{1}{2} n + t - 1 \right) \right] \\
\geq \frac{\exp \left( -nD \left( \frac{1 + n + t + 1}{n} \right) \right)}{\exp \left( \frac{n}{\pi} \sqrt{2\pi} \right)} \cdot \frac{\pi}{\sqrt{2 \left( \frac{1 + n + t - 1}{n} \right) - n} + \sqrt{2\pi}} \\
\geq \frac{\exp \left( -nD \left( \frac{1}{n} + \frac{1}{n} \right) \right)}{\exp \left( \frac{n}{\pi} \sqrt{2\pi} \right)} \cdot \frac{\pi}{\sqrt{2 \left( \frac{1}{n} \right) + \sqrt{2\pi}}} \\
= \exp \left( -\frac{1}{6} \right) \exp \left( -2\psi \left( \frac{t}{n} \right) \right) \cdot \frac{1}{\sqrt{2 \pi} \sqrt{2 \ln \frac{d}{n} + 2} + 2} .
\]

\[
\text{Theorem 7 (Lower Bound on Maximum of Independent Symmetric Random Walks). Let } Z_1^{(n)}, Z_2^{(n)}, \ldots, Z_d^{(n)} \text{ be } d \text{ independent symmetric random walks of length } n. \text{ If } 2 \leq d \leq \exp(\frac{n}{3}) \text{ and } n \geq 7,
\]

\[
E \left[ \max_{1 \leq i \leq d} Z_i^{(n)} \right] \geq 1 - \frac{\exp \left( -\frac{\sqrt{ \ln d} + \sqrt{2\pi} \psi \left( \frac{1}{2} \right) }{\sqrt{\psi \left( \frac{1}{2} \sqrt{\ln d} \right)}} \right) \sqrt{\psi \left( \frac{1}{2} \sqrt{\ln d} \right)} \sqrt{\frac{2 \ln d - 2 \ln \ln d - 1}{\sqrt{\psi \left( \frac{1}{2} \sqrt{\ln d} \right)}}} }{\psi \left( \frac{1}{2} \sqrt{\ln d} \right)} \right] \\
\geq 0.09 \sqrt{n \ln \frac{d}{d} - 2\sqrt{n}} .
\]

Proof. Define the event \( A \) equal to the case that at least one of the \( Z_i^{(n)} \) is greater or equal to \( C \sqrt{n \ln d} - 2 \) where \( C = C(d, n) = \frac{1}{\sqrt{\psi \left( \frac{1}{2} \sqrt{\ln d} \right)}} \sqrt{2 - \frac{2 \ln \ln d}{\ln d}} . \)

We upper and lower bound \( C(d, n) \). Denote by \( f(d) = \sqrt{2 - \frac{2 \ln \ln d}{\ln d}} \) and notice that \( C(d, n) = \frac{1}{\sqrt{\psi \left( \frac{1}{2} \sqrt{\ln d} \right)}} f(d) \).

It suffices to bound \( f(d) \) and \( \psi \left( \frac{1}{2} \sqrt{\ln d} \right) \). We already know that \( 1 \leq \psi(x) \leq 2 \ln(2) \) for all \( x \in [-\frac{1}{2}, \frac{1}{2}] \) and \( \frac{1}{2} \sqrt{\ln d} \in [0, \frac{1}{2}] \) for \( d \leq \exp(n/3) \). The function \( f(d) \) is decreasing on \( (1, e^r] \), increasing on \( [e^r, \infty) \), and \( \lim_{d \to \infty} f(d) = \sqrt{2} \). It has unique minimum at \( e^r \). Therefore, \( f(d) \geq f(e^r) = \sqrt{2 - \frac{2}{e^r}} \geq 1.12 \) for all \( d \in (1, \infty) \). Similarly, from unimodality of \( f(d) \) we have that \( f(d) \leq \max \{ \sqrt{2}, f(2) \} = f(2) \leq 1.6 \) for all \( d \in [2, \infty) \). From this we can conclude that if \( n \geq \ln d > 0, \)

\[
0.95 \leq f(e^r) \sqrt{2 \ln 2} \leq C(d, n) \leq f(2) \leq 1.6 .
\]

If \( n \geq 7 \) and \( 2 \leq d \leq \exp(n/3) \) this implies that

\[
1 < \frac{C \sqrt{n \ln d}}{2} < \frac{n}{2} + 1 .
\]
Recalling the definition of event $A$, we have

\[
\mathbb{E} \left[ \max_{1 \leq i \leq d} Z_1^{(n)} \right] = \mathbb{E} \left[ \max_{1 \leq i \leq d} Z_i^{(n)} \mid A \right] \cdot \Pr[A] + \mathbb{E} \left[ \max_{1 \leq i \leq d} Z_i^{(n)} \mid \overline{A} \right] \cdot \Pr[\overline{A}]
\]

\[
\geq \mathbb{E} \left[ \max_{1 \leq i \leq d} Z_i^{(n)} \mid A \right] \cdot \Pr[A] + \mathbb{E} \left[ Z_i^{(n)} \mid \overline{A} \right] \cdot \Pr[\overline{A}]
\]

\[
\geq \mathbb{E} \left[ \max_{1 \leq i \leq d} Z_i^{(n)} \mid A \right] \cdot \Pr[A] + \mathbb{E} \left[ Z_i^{(n)} \mid Z_i^{(n)} \leq C\sqrt{n \ln d} - 2 \right] \cdot \Pr[\overline{A}]
\]

\[
\geq \mathbb{E} \left[ \max_{1 \leq i \leq d} Z_i^{(n)} \mid A \right] \cdot \Pr[A] + \mathbb{E} \left[ Z_i^{(n)} \mid Z_i^{(n)} \leq 0 \right] \cdot \Pr[\overline{A}]
\]

(by (12))

\[
\geq (C\sqrt{n \ln d} - 2) \Pr[A] + \mathbb{E} \left[ Z_i^{(n)} \mid Z_i^{(n)} \leq 0 \right] \cdot (1 - \Pr[A]).
\]

We lower bound $\mathbb{E} \left[ Z_1^{(n)} \mid Z_1^{(n)} \leq 0 \right]$. Using the fact that distribution of $Z_1^{(n)}$ is symmetric and has zero mean,

\[
\mathbb{E} \left[ Z_1^{(n)} \mid Z_1^{(n)} \leq 0 \right] = \sum_{k=-n}^{0} k \cdot \Pr[Z_1^{(n)} = k \mid Z_1^{(n)} \leq 0]
\]

\[
= \frac{1}{\Pr[Z_1^{(n)} \leq 0]} \sum_{k=-n}^{0} k \cdot \Pr[Z_1^{(n)} = k]
\]

\[
\geq 2 \sum_{k=-n}^{0} k \cdot \Pr[Z_1^{(n)} = k]
\]

(by symmetry of $Z_1^{(n)}$)

\[
= - \sum_{k=-n}^{n} |k| \cdot \Pr[Z_1^{(n)} = k]
\]

(again, by symmetry of $Z_1^{(n)}$)

\[
= - \mathbb{E}[|Z_1^{(n)}|]
\]

\[
= - \mathbb{E} \left[ \sqrt{Z_1^{(n)}}^2 \right]
\]

\[
\geq - \sqrt{\mathbb{E} \left[ Z_1^{(n)} \right]^2}
\]

(by concavity of $\sqrt{\cdot}$)

\[
= - \sqrt{\text{Var} \left( Z_1^{(n)} \right)}
\]

\[
= -\sqrt{n}.
\]

Now let us focus on $\Pr[A]$. Note that $B_n = \frac{Z_1 + n}{2}$ is a binomial random variable with distribution $B(n, \frac{1}{2})$. Similar
to the proof of Theorem 3, we can lower bound $\Pr[A]$ as

$$\Pr[A] = 1 - \Pr[A]$$

$$= 1 - \left( \Pr \left[ Z_1^{(n)} < C\sqrt{n \ln d} - 2 \right] \right)^d$$

$$= 1 - \left( \Pr \left[ B_n < \frac{n}{2} + \frac{C\sqrt{n \ln d}}{2} - 1 \right] \right)^d$$

$$= 1 - \left( 1 - \Pr \left[ B_n \geq \frac{C\sqrt{n \ln d}}{2} + \frac{n}{2} - 1 \right] \right)^d$$

$$\geq 1 - \exp \left( -d \cdot \Pr \left[ B_n \geq \frac{C\sqrt{n \ln d}}{2} + \frac{n}{2} - 1 \right] \right)$$

$$\geq 1 - \exp \left( \frac{- \exp \left( -\frac{1}{6} \right) d^{1 - \frac{C^2}{2}} \psi \left( \frac{C\sqrt{\ln d}}{2\sqrt{n}} \right)}{C\sqrt{2\pi \ln d} + 2} \right)$$

$$\geq 1 - \exp \left( \frac{- \exp \left( -\frac{1}{6} \right) d^{1 - \frac{C^2}{2}} \psi \left( \frac{1.6\sqrt{\ln d}}{2\sqrt{n}} \right)}{1.6\sqrt{2\pi \ln d} + 2} \right)$$

(by Corollary 6 and (12))

We now use the fact that $C = \sqrt{\psi \left( \frac{1.6\sqrt{\ln d}}{2\sqrt{n}} \right)}$ implies that $d^{1 - \frac{C^2}{2}} \psi \left( \frac{1.6\sqrt{\ln d}}{2\sqrt{n}} \right) = \ln d$. Hence, we obtain

$$\Pr[A] \geq 1 - \exp \left( \frac{- \exp \left( -\frac{1}{6} \right) d^{1 - \frac{C^2}{2}} \psi \left( \frac{1.6\sqrt{\ln d}}{2\sqrt{n}} \right)}{1.6\sqrt{2\pi \ln d} + 2} \right)$$

$$= 1 - \exp \left( \frac{- \exp \left( -\frac{1}{6} \right) \ln d}{1.6\sqrt{2\pi \ln d} + 2} \right)$$

$$\geq 1 - \exp \left( \frac{- \exp \left( -\frac{1}{6} \right) \ln d}{2.6\sqrt{2\pi}} \right)$$

$$\geq 1 - \exp \left( \frac{- \ln d}{3.1\sqrt{2\pi}} \right)$$

where in the last equality we used the fact that $\sqrt{2\pi \ln d} > 2$ for $d \geq 2$. Putting all together, we have the stated bound.

\[\square\]

4 Learning with Expert Advice

Learning with Expert Advice is an online problem where in each round $t$ an algorithm chooses (possibly randomly) an action $I_t \in \{1, 2, \ldots, d\}$ and then it receives losses of the actions $\ell_{t,1}, \ell_{t,2}, \ldots, \ell_{t,d} \in [0, 1]$. This repeats for $n$ rounds. The goal of the algorithm is to have a small cumulative loss $\sum_{t=1}^{n} \ell_{t,I_t}$ of actions it has chosen. The difference between the algorithm’s loss and the loss of best fixed action in hind-sight is called regret. Formally,

$$\text{Regret}^{(d)}(n) = \sum_{t=1}^{n} \ell_{t,I_t} - \min_{1 \leq i \leq d} \sum_{t=1}^{n} \ell_{t,i}.$$

There are algorithms that given the number of rounds $n$ as an input achieve regret no more than $\sqrt{\frac{n}{2} \ln d}$ for any sequence of losses.
Theorem 8. Let \( n \geq 7 \) and \( 2 \leq d \leq \exp(\frac{n}{3}) \). For any algorithm for learning with expert advice there exists a sequence of losses \( \ell_{t,i} \in \{0, 1\} \), \( 1 \leq i \leq d \), \( 1 \leq t \leq n \), such that

\[
\operatorname{Regret}^{(d)}(n) \geq \frac{1}{\sqrt{n}} \left( \sqrt{2 \ln d - 2 \ln \ln d - 1} - \frac{\sqrt{n}}{2} \right).
\]

Proof. Proceeding as in the proof of Theorem 3.7 in [Cesa-Bianchi and Lugosi, 2006] we only need to show that

\[
\operatorname{Regret}^{(d)}(n) \geq \frac{1}{2} \mathbb{E} \left[ \max_{1 \leq i \leq d} Z_{i}^{(n)} \right]
\]

where \( Z_{1}^{(n)}, Z_{2}^{(n)}, \ldots, Z_{d}^{(n)} \) are independent symmetric random walks of length \( n \). The theorem follows from Theorem 7.

The theorem proves a non-asymptotic lower bounds, while at the same time recovering the optimal constant of the asymptotic one in [Cesa-Bianchi and Lugosi, 2006].

References

A. V. Boyd. Inequalities for Mill’s ratio. Reports of Statistical Application Research (Union of Japanese Scientists and Engineers), 6:44–46, 1959.

N. Cesa-Bianchi and G. Lugosi. Prediction, learning, and games. Cambridge University Press, 2006.

B. D. McKay. On Littlewood's estimate for the binomial distribution. Advanced Applied Probability, 21:475–478, 1989.

F. Orabona. Dimension-free exponentiated gradient. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, Advances in Neural Information Processing Systems 26, pages 1806–1814. Curran Associates, Inc., 2013.

H. Robbins. A remark on Stirling’s formula. American Mathematical Monthly, pages 26–29, 1955.

A Upper Bounds

We say that a random variable \( X \) is \( \sigma^{2} \)-sub-Gaussian (for some \( \sigma \geq 0 \)) if

\[
\mathbb{E} \left[ e^{sX} \right] \leq \exp \left( \frac{\sigma^{2}s^{2}}{2} \right) \quad \text{for all } s \in \mathbb{R}.
\]

(13)

It is straightforward to verify that \( X \sim N(0, \sigma^{2}) \) is \( \sigma^{2} \)-sub-Gaussian. Indeed, for any \( s \in \mathbb{R} \),

\[
\mathbb{E} \left[ e^{sX} \right] = \int_{-\infty}^{\infty} \frac{1}{\sigma \sqrt{2\pi}} \exp \left( -\frac{x^{2}}{2\sigma^{2}} \right) e^{sx} dx
\]

\[
= \exp \left( \frac{s^{2}\sigma^{2}}{2} \right) \int_{-\infty}^{\infty} \frac{1}{\sigma \sqrt{2\pi}} \exp \left( -\frac{(x - s\sigma)^{2}}{2\sigma^{2}} \right) dx
\]

\[
= \exp \left( \frac{s^{2}\sigma^{2}}{2} \right).
\]
We now show that a Rademacher random variable $Y$ (with distribution $\Pr[Y = +1] = \Pr[Y = -1] = \frac{1}{2}$) is $1$-sub-Gaussian. Indeed, for any $s \in \mathbb{R}$,

$$\mathbb{E}[e^{sY}] = \frac{e^s + e^{-s}}{2} = \frac{1}{2} \sum_{k=0}^{\infty} \frac{s^k}{k!} + \frac{1}{2} \sum_{k=0}^{\infty} (-1)^k \frac{s^k}{k!} = \sum_{k=0}^{\infty} \frac{2k^k}{(2k)!} \leq \sum_{k=0}^{\infty} \frac{e^{2k^k}}{k!^2} = \exp \left( \frac{s^2}{2} \right).$$

If $Y_1, Y_2, \ldots, Y_n$ are independent $\sigma$-sub-Gaussian random variables, then $\sum_{i=1}^{n} Y_i$ is $(n\sigma^2)$-sub-Gaussian. This follows from

$$\mathbb{E}[e^{s\sum_{i=1}^{n} Y_i}] = \prod_{i=1}^{n} \mathbb{E}[e^{sY_i}].$$

This property proves that the symmetric random walk $Z^{(n)}$ of length $n$ is $n$-sub-Gaussian.

The upper bounds (1) and (3) follow directly from sub-Gaussianity of the variables involved and the following lemma.

**Lemma 9** (Maximum of sub-Gaussian random variables). Let $X_1, X_2, \ldots, X_d$ be (possibly dependent) $\sigma^2$-sub-Gaussian condition random variables. Then,

$$\mathbb{E}\left[ \max_{1 \leq i \leq d} X_i \right] \leq \sigma \sqrt{2 \ln d}.$$

**Proof.** For any $s > 0$, we have

$$\mathbb{E}\left[ \max_{1 \leq i \leq d} X_i \right] = \frac{1}{s} \mathbb{E}\left[ \max_{1 \leq i \leq d} \ln e^{sX_i} \right] \leq \frac{1}{s} \ln \mathbb{E}\left[ \max_{1 \leq i \leq d} e^{sX_i} \right] \leq \frac{1}{s} \ln \mathbb{E}\left[ \sum_{i=1}^{d} e^{sX_i} \right] = \frac{1}{s} \ln \sum_{i=1}^{d} \mathbb{E}[e^{sX_i}] \leq \frac{1}{s} \ln \left( d \exp \left( \frac{\sigma^2 s^2}{2} \right) \right) = \frac{\ln d}{s} + \frac{\sigma^2 s^2}{2}.$$

Substituting $s = \sqrt{2 \ln d}$ finishes the proof. \qed