The Best Expert Versus the Smartest Algorithm

Peter Chen
Department of Computer Science,
Louisiana State University,
Baton Rouge, Louisiana, 70803, USA

Guoli Ding
Department of Mathematics,
Louisiana State University,
Baton Rouge, Louisiana, 70803, USA

June 27, 2003

Abstract

In this paper, we consider the problem of *online prediction using expert advice*. Under different assumptions, we give tight lower bounds on the gap between the best expert and any online algorithm that solves the problem.

Key words. Online algorithm, online prediction, expert advice.

1 Introduction

The problem of *online prediction using expert advice* is for a predictor to predict, along with n other "experts", a sequence $\sigma = \sigma_1, \sigma_2, ..., \sigma_\ell \in \{0, 1\}^\ell$. Here, the only assumptions we make are the following.

Assumption 1.1 *Before predicting each σ_j, the predictor knows the predictions of the experts on this term. Also, right after predicting each σ_j, the predictor is given the true value of this term.*

Notice that we do not assume anything on possible patterns of either σ or the sequences of the predictions of the experts. The goal of the predictor is to "score" as close to the best expert as possible. We point out that this is different from the goal that tries to predict as accurate as possible, which is a problem studied in [1].

Next, we make the problem more precise. Suppose $x = x_1, x_2, ..., x_\ell$ is a sequence of predictions, made by the predictor or by the experts. It is worth mentioning that, sometimes, terms of x might be allowed to take any value in the interval $[0, 1]$. The *loss* of x is defined to be

$$L(x, \sigma) = \sum_{j=1}^{\ell} |x_j - \sigma_j|.$$

Let γ_i be the sequence of predictions made by expert i and let $\Gamma = \{\gamma_i : 1 \le i \le n\}$. Then

$$L(\Gamma, \sigma) = \min\{L(\gamma_i, \sigma) : 1 \le i \le n\}$$

is the loss of the best expert. For any strategy $\mathcal{A}$ of the predictor,[1] let $\tau_{\mathcal{A}}(\sigma, \Gamma)$ be the sequence of predictions generated according to $\mathcal{A}$. We measure the performance of $\mathcal{A}$ by the worst *gap* between the losses of the predictor and the best expert. That is, by

$$G_{\mathcal{A}}(n, \ell) = \sup_{\sigma, \Gamma} \; (L(\tau_{\mathcal{A}}(\sigma, \Gamma), \sigma) - L(\Gamma, \sigma)).$$

[1] Here we assume that, when predicting two sequences σ' and σ'', if σ' and σ'' turn out to be the same, and the predictions of the experts are also the same, then the strategy should generate two identical sequences of predictions. For a strategy that generates two different sequences of predictions, due to randomization or similar reasons, we will consider it as several strategies (under our term) and this change does not affect our discussions.

Clearly, the goal of the predictor is to minimize $G_{\mathcal{A}}(n, \ell)$ over all strategies $\mathcal{A}$. In this paper, we analyze upper and lower bounds of $G_{\mathcal{A}}(n, \ell)$.

To get a lower bound, let us make the following assumption, which is more in favor of the predictor, and thus will make the result stronger:

Assumption 1.2 *The predictions of the predictor can be any real number in the interval* $[0, 1]$, *while the predictions of the experts can only be* 0 *or* 1. *In addition, before predicting* σ_1, *the predictor knows not only* ℓ, *but also* Γ, *the entire prediction sequence of each expert. In other words, other than the actual value of each* σ_j, *the predictor knows everything else before predicting* σ_1.

Under Assumption 1.2, it is proved in [2] that, for all strategies $\mathcal{A}$,

$$\liminf_{n\to\infty} \; \liminf_{\ell\to\infty} \frac{G_{\mathcal{A}}(n, \ell)}{\sqrt{(\ell/2)\ln n}} \geq 1. \tag{1.1}$$

We should point out that, because of the order the two limits are taken, it is assumed, implicitly, in the above inequality that ℓ is significantly larger than n. We will see later that the situation is quite different if ℓ is smaller than n.

For upper bounds, let us make a different assumption, which is less in favor of the predictor, and thus will make the result stronger.

Assumption 1.3 *The predictions of the predictor and the experts can be any real number in* $[0, 1]$. *The predictor also knows* ℓ *before predicting* σ_1.

Under Assumption 1.3, an online algorithm (a strategy for the predictor) $\mathcal{A}$ is given in [2] for which

$$\liminf_{n\to\infty} \; \liminf_{\ell\to\infty} \frac{G_{\mathcal{A}}(n, \ell)}{\sqrt{(\ell/2)\ln n}} \leq 1. \tag{1.2}$$

In fact, what has been proved is that, for all positive integers n and ℓ, the algorithm $\mathcal{A}$ satisfies

$$G_{\mathcal{A}}(n, \ell) \leq \sqrt{\frac{\ell \ln(n+1)}{2}} + \frac{\log_2(n+1)}{2}. \tag{1.3}$$

Notice that (1.3) is much stronger than (1.2) since it upper bounds $G_{\mathcal{A}}(n, \ell)$ for all n and ℓ. Having such a bound is important because very often, in various applications, n and ℓ are not arbitrarily large. In this paper, we improve lower bound (1.1) in the same way.

Theorem 1.1 *Under Assumption 1.2, for any algorithm* $\mathcal{A}$, *any integer* $n \geq 2$, *and any* $\epsilon \in [0, 1]$, *if*

$$\ell \geq \ell(n) := \sqrt{\pi/8}\ ((\ln n)^2 + 8)\ (\sqrt{2\ln n} + 1)\ n^{1-\epsilon}, \tag{1.4}$$

then

$$G_{\mathcal{A}}(n, \ell) \geq \left(\sqrt{\frac{(1-\epsilon)\ell \ln n}{2}} - \frac{1}{2} \right) \ (1 - \delta^n - (1-\delta)^n), \tag{1.5}$$

where

$$\delta = \frac{1}{\sqrt{2\pi}\ (\sqrt{2\ln n} + 1)\ n^{1-\epsilon}} - \frac{(\ln n)^2 + 8}{4\ell}.$$

First, as easily shown below, (1.1) is a consequence of Theorem 1.1, and thus our theorem is indeed an improvement of (1.1) (in the sense that our result implies (1.1) yet it is not implied by (1.1)).

Corollary 1.1 *Inequality (1.1) holds for all online prediction algorithms* $\mathcal{A}$.

Proof. By setting $\delta_1 = 1/(\sqrt{2\pi}(1+\sqrt{2\ln n}\,)n^{1-\epsilon})$, it is straightforward to verify that

$$\begin{aligned}\liminf_{n\to\infty}\ \liminf_{\ell\to\infty} \frac{G_{\mathcal{A}}(n,\ell)}{\sqrt{(\ell/2)\ln n}} &\geq \liminf_{n\to\infty}\ \liminf_{\ell\to\infty}\left(\sqrt{1-\epsilon}-\frac{1}{\sqrt{2\ell\ln n}}\right)(1-\delta^n-(1-\delta)^n)\\ &= \liminf_{n\to\infty}\sqrt{1-\epsilon}(1-\delta_1^n-(1-\delta_1)^n)\\ &= \sqrt{1-\epsilon},\end{aligned}$$

holds for all $\epsilon \in (0,1)$, and so (1.1) follows. ∎

Further remarks on Theorem 1.1.

(a) If $n = 1$, it is easy to see that the algorithm that copies the only expert will perform exactly the same as the best expert and thus $G_{\mathcal{A}}(n,\ell) = 0$, for all ℓ. Because of this, we can say that the assumption $n \geq 2$ in the theorem does not miss any interesting cases.

(b) Inequality (1.5) still holds if we set $\epsilon = 0$. We introduce this extra parameter because we need it in proving Corollary 1.1.

(c) For any $\epsilon > 0$, it is clear that $\ell(n)/n \longrightarrow 0$, as $n \longrightarrow \infty$. Therefore, the requirement $\ell \geq \ell(n)$ is more or less the same as $\ell \geq n$.

(d) One may wonder if the requirement $\ell \geq \ell(n)$ can be dropped completely. For instance, one may ask if there could exist a constant $c > 0$, a function $d(n,\ell)$ with $\lim_{n\to\infty} d(n,\ell) = 0$, and such that

$$G_{\mathcal{A}}(n,\ell) \geq \sqrt{\frac{\ell \ln n}{2}}\,(c + d(n,\ell)) \tag{1.6}$$

holds for all n, ℓ, and $\mathcal{A}$. Unfortunately, the answer is negative. Consider the strategy $\mathcal{A}$ that predicts 1/2 all the time. Then $L(\tau_{\mathcal{A}}(\sigma,\Gamma),\sigma) = \ell/2$, for all σ and Γ. It follows that

$$G_{\mathcal{A}}(n,\ell) = \sup_{\sigma,\Gamma}\ (L(\tau_{\mathcal{A}}(\sigma,\Gamma),\sigma) - L(\Gamma,\sigma)) = \frac{\ell}{2} - \inf_{\sigma,\Gamma}\ L(\Gamma,\sigma) \leq \frac{\ell}{2}. \tag{1.7}$$

Clearly, this inequality contradicts (1.6), for every $\ell > 0$, when n is sufficiently large. This contradiction indicates that a condition similar to $\ell \geq \ell(n)$ is required to prove any lower bound of the form (1.6).

So far we have discussed the situation when ℓ is bigger than n. When ℓ is smaller than n, we have seen from Remark (d) that lower bounds (1.1) and (1.5) no longer hold. Moreover, as indicated by our next result, that upper bounds (1.2) and (1.3) are not very close to the truth either.

Theorem 1.2 *For all n, ℓ, and $\mathcal{A}$, under Assumption 1.2, we have*

$$G_{\mathcal{A}}(n,\ell) \geq \frac{\ell}{2}(1-(1-2^{-\ell})^n-(2^{-\ell})^n).$$

By combining this result with (1.7) we obviously have the following.

Corollary 1.2 *For all ℓ, under Assumption 1.2,*

$$\lim_{n\to\infty}\ \inf_{\mathcal{A}}\ G_{\mathcal{A}}(n,\ell) = \frac{\ell}{2}.$$

This result suggests that, if n is significantly larger than ℓ, then the predictor cannot catch up with the best expert. The only thing the predictor can do is to predict 1/2 all the time so that it won't be left too far behind the best expert.

Remaining questions. We have seen the behavior of $G_{\mathcal{A}}(n,\ell)$ when n and ℓ are far apart. But the situation is not very clear when n and ℓ are very close. For instance, it would be very interesting to know the asymptotic behavior of $\inf_{\mathcal{A}} G_{\mathcal{A}}(2^\ell,\ell)$, or the exactly value of $\inf_{\mathcal{A}} G_{\mathcal{A}}(n,n)$.

We closed this section by giving an outline of the rest of the paper. In Section 2, we present all necessary mathematical formulas. Then we prove Theorem 1.2 in Section 3. The proof of Theorem 1.1 will be given in the last section.

2 Preliminaries

In this section, we present all necessary mathematical tools needed for the rest of the paper. Our first lemma is well known and can be found in many Calculus books.

Lemma 2.1 (Stirling Formula). *For each positive integer m,*

$$m! = \sqrt{2\pi m}\left(\frac{m}{e}\right)^m e^{\theta_m},$$

where θ_m satisfies

$$0 < \theta_m < \frac{1}{12m}.$$

The next lemma consists of some basic facts about normal distribution, which can also be found in most Probability books.

Lemma 2.2 *Let $\Phi(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-x^2/2}dx$. Then*

(i) $\Phi(0) = 1/2$;

(ii) $\Phi(-z) = 1 - \Phi(z)$, for all $z > 0$; and

(iii) $\Phi(-z) \geq \frac{1}{\sqrt{2\pi}\,(z+1)} e^{-z^2/2}$, for all $z > 0$.

Lemma 2.3 *If $n > 1$, then $\alpha(x) = x^n + (1-x)^n$ is decreasing over the interval $[0,\ 1/2]$.*

Proof. Since $n > 1$ and $x \in (0,\ 1/2)$, it is clear that $\alpha'(x) = n(x^{n-1} - (1-x)^{n-1}) < 0$, and thus the lemma follows. ■

Lemma 2.4 *For every x in $[0,\ 1/2]$, we have*

(i) $\ln(1+x) \geq x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4}$, and

(ii) $\ln(1-x) \geq -x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{2}$.

Proof. Notice that both inequalities are equalities when $x = 0$. In addition,

$$\left(\ln(1+x) - x + \frac{x^2}{2} - \frac{x^3}{3} + \frac{x^4}{4}\right)' = \frac{1}{1+x} - 1 + x - x^2 + x^3 = \frac{x^4}{1+x} > 0$$

holds for all $x > 0$, and thus (i) follows. Similarly,

$$\left(\ln(1-x) + x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{2}\right)' = \frac{-1}{1-x} + 1 + x + x^2 + 2x^3 = \frac{x^3(1-2x)}{1-x} > 0$$

holds for all $x \in (0, 1/2)$, and thus (ii) follows. ■

Lemma 2.5 *Let x and y be nonnegative numbers. Then $e^{-x-y} - e^{-y} \geq -x$.*

Proof. Notice that the inequality is an equality when $x = 0$. Moreover, when taking derivative with respect to x we have $(e^{-x-y} - e^{-y} + x)' = -e^{-x-y} + 1 \geq 0$. Therefore, the result follows. ■

Lemma 2.6 *Let $f(x)$ be a function defined on the interval $[c, c+2]$.*

(i) If $f''(x) > 0$ on $(c, c+2)$, then $\int_c^{c+2} (f(x) - f(c+1))dx \geq 0$;

(ii) If $f''(x) < 0$ on $(c, c+2)$, then $\int_c^{c+1} (f(x) - f(c))dx + \int_{c+1}^{c+2} (f(x) - f(c+2))dx \geq 0$.

Proof. If $f''(x) > 0$, then $f(x)$ is convex and thus $\frac{1}{2}f(x_1) + \frac{1}{2}f(x_2) \geq f(\frac{x_1+x_2}{2})$ holds for all x_1, x_2 in $[c, c+2]$. Consequently,

$$\begin{aligned}
\int_c^{c+2} (f(x) - f(c+1))dx &= \int_c^{c+1} (f(x) - f(c+1))dx + \int_{c+1}^{c+2} (f(x) - f(c+1))dx \\
&= \int_c^{c+1} (f(x) - f(c+1))dx + \int_c^{c+1} (f(2c+2-x) - f(c+1))dx \\
&= \int_c^{c+1} (f(x) + f(2c+2-x) - 2f(c+1))dx \\
&\geq 0,
\end{aligned}$$

as required. On the other hand, if $f''(x) < 0$, then $f(x)$ is concave and thus, for all $x_1, x_2 \in [c, c+2]$ and all $\lambda \in [0,1]$, we have $f(\lambda x_1 + (1-\lambda)x_2) \geq \lambda f(x_1) + (1-\lambda)f(x_2)$ and $f((1-\lambda)x_1 + \lambda x_2) \geq (1-\lambda)f(x_1) + \lambda f(x_2)$. Let $x = \lambda x_1 + (1-\lambda)x_2$. Then adding the last two inequalities gives us $f(x) + f(x_1 + x_2 - x) \geq f(x_1) + f(x_2)$. Now it is clear that

$$\begin{aligned}
&\int_c^{c+1} (f(x) - f(c))dx + \int_{c+1}^{c+2} (f(x) - f(c+2))dx \\
= &\int_c^{c+1} (f(x) - f(c))dx + \int_c^{c+1} (f(2c+2-x) - f(c+2))dx \\
\geq &\ 0,
\end{aligned}$$

and thus the lemma is proved. ■

Lemma 2.7 *The inequality $\frac{z^4+2}{\Phi(-2z)} \geq \max\{50z^2,\ 4(8z^4+1)\}$ holds for all $z > 0$.*

Proof. It is easy to see that, for $z > 0$,

$$\max\{50z^2,\ 4(8z^4+1)\} = \begin{cases} 4(8z^4+1) & 0 < z \leq \alpha_1 \\ 50z^2 & \alpha_1 \leq z \leq \alpha_2 \\ 4(8z^4+1) & \alpha_2 \leq z < \infty \end{cases}$$

where $\alpha_1 = \sqrt{(25 - \sqrt{497})/32} \approx 0.2908$ and $\alpha_2 = \sqrt{(25 + \sqrt{497})/32} \approx 1.2157$ are the only two positive solutions to the equation $50z^2 = 4(8z^4+1)$. Therefore, to prove the Lemma, we only need to prove the following three claims.

Claim 1. $z^4 + 2 \geq 4(8z^4+1)\Phi(-2z)$, for $z \in (1, \infty)$.
Claim 2. $z^4 + 2 \geq 4(8z^4+1)\Phi(-2z)$, for $z \in (0, 0.3)$.
Claim 3. $z^4 + 2 \geq 50z^2\Phi(-2z)$, for $z \in (0, \infty)$.

Since $\Phi(z)$ is an increasing function and $\Phi(-2) \approx 0.02275$, it follows that

$$4(8z^4+1)\Phi(-2z) \le 4(8z^4+1)\cdot 0.025 = 0.8z^4 + 0.1 < z^4 + 2,$$

and thus Claim 1 is proved. To prove Claim 2, let

$$f(z) = \frac{z^4+2}{4(8z^4+1)} - \Phi(-2z).$$

Then

$$f'(z) = \frac{\sqrt{\frac{2}{\pi}}\ (1+8z^4)^2 - 15z^3 e^{2z^2}}{(1+8z^4)^2 e^{2z^2}},$$

and

$$f''(z) = -\frac{z\left(4\sqrt{\frac{2}{\pi}}\ (1+8z^4)^3 + 15z(3-40z^4)e^{2z^2}\right)}{(1+8z^4)^3 e^{2z^2}}.$$

For each $z \in (0, 0.3)$, it is easy to see that $3 - 40z^4 > 0$, and so we have $f''(z) \le 0$. Consequently, $f'(z)$ is a decreasing function, which implies that

$$f(z) \ge \min\{f(0), f(0.3)\} = f(0) = 0,$$

for all $z \in (0, 0.3)$, which proves Claim 2.

We prove Claim 3 by considering three intervals, $(0, 0.73)$, $[0.73, 0.8]$, and $(0.8, \infty)$. If $z \in (0.8, \infty)$, we deduce from $\Phi(-2\cdot 0.8) \approx 0.0548 < 0.056$ that

$$z^4 + 2 - 50z^2\Phi(-2z) > z^4 + 2 - 2.8z^2 = (z^2 - 1.4)^2 + 0.4 > 0,$$

and thus the desired inequality holds. Next, let

$$g(z) = \frac{z^4+2}{50z^2}.$$

Then $g(z)$ is a decreasing function, over $[0.73, 0.8]$, since $g'(z) = -0.04(2 - z^4)/z^3$. Recall that $\Phi(z)$ is an increasing function. Therefore, for $z \in [0.73, 0.8]$,

$$g(z) \ge g(0.8) = 0.0753 > 0.0721 \approx \Phi(-2\cdot 0.73) \ge \Phi(-2z),$$

which is what we want.

It remains to prove $z^4 + 2 \ge 50z^2\Phi(-2z)$, over $(0, 0.73)$. In the remainder of this proof, we assume that z belongs to $(0, 0.73)$. Let

$$h_1(z) = z^2 + 2z^{-2} - 50\Phi(-2z) \quad \text{and} \quad h_2(z) = 2z^2 - 5\ln z.$$

Then

$$h_1'(z) = 2z - 4z^{-3} + \frac{100}{\sqrt{2\pi}}e^{-2z^2}, \quad h_1''(z) = 2 + 12z^{-4} - \frac{400z}{\sqrt{2\pi}}e^{-2z^2},$$

and

$$h_2'(z) = -\frac{5-4z^2}{z}.$$

Clearly, $h_2'(z) < 0$, which implies that $h_2(z)$ is a decreasing function. Consequently,

$$h_2(z) \ge h_2(0.73) \approx 2.63935 > 2.58762 \approx \ln\frac{400}{12\sqrt{2\pi}},$$

which implies

$$2z^2 > \ln \frac{400z^5}{12\sqrt{2\pi}}.$$

Therefore,

$$h_1''(z) > 12z^{-4} - \frac{400z}{\sqrt{2\pi}} e^{-2z^2} = \frac{12}{z^4 e^{2z^2}} \left(e^{2z^2} - \frac{400z^5}{12\sqrt{2\pi}} \right) > 0,$$

which implies that $h_1'(z)$ is an increasing function. Notice that $h_1'(0.5) \approx -6.8 < 0$ and $h_1'(0.6) \approx 2.1 > 0$. It follows that $h_1'(z) = 0$ has a unique solution, say z_0, which means that $h_1(z)$ achieves its minimum at z_0. To complete our proof, we only need to show that $h_1(z_0) \geq 0$. For this, we apply Taylor's Expansion Theorem to $h_1(z)$ at $z = 0.56$. We have

$$h_1(z) = h_1(0.56) + h_1'(0.56)(z - 0.56) + \frac{1}{2} h_1''(\xi)(z - 0.56)^2,$$

where ξ is a number between z and 0.56. Since $h_1'(0.56) \approx -0.35 < 0$, we conclude that z_0 is between 0.56 and 0.6. Now, from $h_1''(z) > 0$ and $h_1(0.56) \approx 0.123$ we deduce that

$$h_1(z_0) \geq h_1(0.56) + h_1'(0.56)(z_0 - 0.56) \geq 0.116 - 0.4(0.6 - 0.56) = 0.1 > 0,$$

which competes our proof of the Lemma. ∎

The last is another technical lemma, which we will use in the last section.

Lemma 2.8 *Let λ and μ be positive numbers and let $g(x) = \frac{e^{\lambda x}}{\sqrt{1 - \mu x}} - 1 - 1.16(\lambda + \frac{\mu}{2})x$. Then $g(x) \leq 0$ if $0 \leq x \leq \min\{\frac{1}{48\lambda}, \frac{2}{25\mu}\}$.*

Proof. Notice that $g(0) = 0$ and

$$g'(x) = \frac{e^{\lambda x}}{(1 - \mu x)^{3/2}} (\lambda(1 - \mu x) + \frac{\mu}{2}) - 1.16(\lambda + \frac{\mu}{2}) \leq \frac{e^{1/48}}{(0.92)^{3/2}} (\lambda + \frac{\mu}{2}) - 1.16(\lambda + \frac{\mu}{2}) \leq 0.$$

Thus the result follows. ∎

3 Proving Theorem 1.2

It is clear from the definition of $G_{\mathcal{A}}(n, \ell)$ that, in order to prove a lower bound, we only need to find one sequence σ and one set Γ of sequences, for every $\mathcal{A}$, such that $L(\tau_{\mathcal{A}}(\sigma, \Gamma), \sigma) - L(\Gamma, \sigma)$ is greater than or equal to the lower bound. Since there are too many choices for $\mathcal{A}$, it is difficult (or may be impossible) to choose σ and Γ. The way to solve this problem is to use probabilistic method.

Recall that, under Assumption 1.2, the predictions of the experts must be 0 or 1. Suppose each expert makes prediction by tossing, independently, a fair coin. In addition, suppose the outcome σ is also determined by tossing a fair coin. In the following, we establish a lower bound on the expected gap between the losses of any $\mathcal{A}$ and the best expert. Then we prove Theorem 1.2 using this bound.

For each $i = 1, 2, ..., n$ and $j = 1, 2, ..., \ell$, let $X_{i,j}$ be the random variable such that $X_{i,j} = 0$, if the prediction of expert i on σ_j is correct, and $X_{i,j} = 1$, if this prediction is wrong. Then it is not difficult to verify that $\mathcal{X} = \{X_{i,j} : \ 1 \leq i \leq n \ \text{ and } \ 1 \leq j \leq \ell\}$ is a set of mutually independent random variables with $P(x_{i,j} = 0) = P(X_{i,j} = 1) = 1/2$. For each $i = 1, 2..., n$, let $S_i = X_{i,1} + X_{i,2} + ... + X_{i,\ell}$, which is the number of mistakes made by expert i. It follows that the number of mistakes made by the best expert is $S = \min\{S_1, S_2, ..., S_n\}$.

Lemma 3.1 *For all $j = 0, 1, ..., \ell$, let $p_j = p_j(\ell) = 2^{-\ell}\binom{\ell}{j}$. Then*

$$E[S] = \sum_{i=1}^{\ell} \left(\sum_{j=i}^{\ell} p_j \right)^n .$$

Proof. If x is a nonnegative integer not exceeding ℓ, it is clear that

$$P(S \geq x) = P(S_1 \geq x; S_2 \geq x; ...; S_n \geq x) = \prod_{i=1}^{n} P(S_i \geq x) = \left(\sum_{j=x}^{\ell} p_j \right)^n ,$$

and so,

$$P(S = x) = P(S \geq x) - P(S \geq x+1) = \left(\sum_{j=x}^{\ell} p_j \right)^n - \left(\sum_{j=x+1}^{\ell} p_j \right)^n .$$

Therefore,

$$\begin{aligned}
E[S] &= \sum_{x=0}^{\ell} x \left(\left(\sum_{j=x}^{\ell} p_j \right)^n - \left(\sum_{j=x+1}^{\ell} p_j \right)^n \right) \\
&= \sum_{x=1}^{\ell} x \left(\sum_{j=x}^{\ell} p_j \right)^n - \sum_{x=0}^{\ell-1} x \left(\sum_{j=x+1}^{\ell} p_j \right)^n \\
&= \sum_{x=1}^{\ell} x \left(\sum_{j=x}^{\ell} p_j \right)^n - \sum_{x=1}^{\ell} (x-1) \left(\sum_{j=x}^{\ell} p_j \right)^n \\
&= \sum_{x=1}^{\ell} \left(\sum_{j=x}^{\ell} p_j \right)^n ,
\end{aligned}$$

and the lemma is proved. ■

Lemma 3.2 *Under Assumption 1.2, for all n, ℓ, and $\mathcal{A}$, we have $G_{\mathcal{A}}(n,\ell) \geq G(n,\ell)$, where*

$$G(n,\ell) = \frac{\ell}{2} - \sum_{i=1}^{\ell} \left(\sum_{j=i}^{\ell} p_j \right)^n . \tag{3.1}$$

Proof. For each $j = 1, 2, ..., \ell$, let Y_j be the outcome of σ_j. Here, we assume that $\mathcal{Y} = \{Y_j : j = 1, 2, ..., \ell\}$ is a set of mutually independent random variables with $P(Y_j = 0) = P(Y_j = 1) = 1/2$. For any algorithm $\mathcal{A}$, let $\tau_{\mathcal{A}}(\mathcal{Y}, \mathcal{X}) = \tau_1, \tau_2, ..., \tau_\ell$. Notice that, for each j,

$$E[|\tau_j - Y_j|] = \frac{\tau_j}{2} + \frac{1-\tau_j}{2} = \frac{1}{2}.$$

Therefore, $E[L(\tau_{\mathcal{A}}(\mathcal{Y}, \mathcal{X}), \mathcal{Y})] = \ell/2$. Also notice that $E[L(\mathcal{X}, \mathcal{Y})] = E(S)$, so, from Lemma 3.1, we deduce that $E[L(\tau_{\mathcal{A}}(\mathcal{Y}, \mathcal{X}), \mathcal{Y}) - L(\mathcal{X}, \mathcal{Y})] = G(n,\ell)$. Consequently, for some special values of $\mathcal{Y}$ and $\mathcal{X}$, say σ and Γ, we must have $L(\tau_{\mathcal{A}}(\sigma, \Gamma), \sigma) - L(\Gamma, \sigma) \geq G(n,\ell)$, and thus the lemma follows. ■.

For each $i = 1, 2, ..., \ell+1$, let

$$P_i = P_i(\ell) = \sum_{j=0}^{i-1} p_j(\ell). \tag{3.2}$$

Lemma 3.3 *For all n and ℓ, we have* $G(n,\ell) = \frac{\ell}{2} - \sum_{i=1}^{\lfloor \ell/2 \rfloor}((1-P_i)^n + P_i^n) - \Delta(n,\ell)$, *where*

$$\Delta(n,\ell) = \begin{cases} 0 & \text{if } \ell \text{ is even,} \\ 1/2^n & \text{if } \ell \text{ is odd.} \end{cases}$$

Proof. From (3.1) it is clear that

$$\begin{aligned} G(n,\ell) &= \frac{\ell}{2} - \sum_{i=1}^{\lfloor \ell/2 \rfloor} \left(\sum_{j=i}^{\ell} p_j \right)^n - \sum_{i=\lceil \ell/2 \rceil + 1}^{\ell} \left(\sum_{j=i}^{\ell} p_j \right)^n - \Delta(n,\ell) \\ &= \frac{\ell}{2} - \sum_{i=1}^{\lfloor \ell/2 \rfloor} \left(1 - \sum_{j=0}^{i-1} p_j \right)^n - \sum_{i=\lceil \ell/2 \rceil + 1}^{\ell} \left(\sum_{j=i}^{\ell} p_j \right)^n - \Delta(n,\ell). \end{aligned} \tag{3.3}$$

Notice that, by setting $i' = \ell + 1 - i$, $j' = \ell - j$, and using $p_k = p_{\ell-k}$, we have

$$\sum_{i=\lceil \ell/2 \rceil + 1}^{\ell} \left(\sum_{j=i}^{\ell} p_j \right)^n = \sum_{i'=1}^{\lfloor \ell/2 \rfloor} \left(\sum_{j=\ell+1-i'}^{\ell} p_j \right)^n = \sum_{i'=1}^{\lfloor \ell/2 \rfloor} \left(\sum_{j'=0}^{i'-1} p_{\ell-j'} \right)^n = \sum_{i=1}^{\lfloor \ell/2 \rfloor} \left(\sum_{j=0}^{i-1} p_j \right)^n,$$

and thus the lemma follows from (3.3) and (3.2). ■

Proof of Theorem 1.2. In ℓ is even, by Lemma 3.2, Lemma 3.3, and Lemma 2.3, we have

$$G_{\mathcal{A}}(n,\ell) \geq \frac{\ell}{2} - \sum_{i=1}^{\ell/2}((1-P_1)^n + P_1^n) = \frac{\ell}{2} - \frac{\ell}{2}((1-\frac{1}{2^\ell})^n + (\frac{1}{2^\ell})^n)$$

and the result follows. If ℓ is odd, a little extra effort is needed. When $\ell = 1$, it is straightforward to verify from (3.1) that

$$G(n,\ell) = \frac{1}{2} - \frac{1}{2^n} = \frac{1}{2}(1 - (1-\frac{1}{2})^n - (\frac{1}{2})^n),$$

and thus the theorem holds in this case. When $\ell \geq 3$, again, using Lemma 3.2, Lemma 3.3, and Lemma 2.3, we have

$$\begin{aligned} G_{\mathcal{A}}(n,\ell) &\geq \frac{\ell}{2} - \sum_{i=1}^{(\ell-1)/2} ((1-P_1)^n + P_1^n) - \Delta(n,\ell) \\ &= \frac{\ell}{2} - \frac{\ell-1}{2}((1-\frac{1}{2^\ell})^n + (\frac{1}{2^\ell})^n) - \frac{1}{2^n} \\ &\geq \frac{\ell}{2}(1-(1-\frac{1}{2^\ell})^n - (\frac{1}{2^\ell})^n) + \frac{1}{2}((1-\frac{1}{8})^n + (\frac{1}{8})^n) - (\frac{1}{2})^n \\ &\geq \frac{\ell}{2}(1-(1-\frac{1}{2^\ell})^n - (\frac{1}{2^\ell})^n). \end{aligned}$$

The theorem is proved. ■

4 Proving Theorem 1.1

We prove the theorem by proving a sequence of lemmas. Our starting point is Lemma 3.2. In the rest of the paper, let z be a positive number and let

$$a = \left\lceil \frac{\ell-1}{2} - z\sqrt{\ell} \right\rceil. \tag{4.1}$$

Lemma 4.1 *For every positive integer ℓ, we have $G(n,\ell) \geq (\frac{\ell}{2} - a)\ (1 - \alpha(P_{a+1}))$, where α is the function defined in Lemma 2.3.*

Proof. It follows from Lemma 3.3 and Lemma 2.3 that

$$\begin{aligned} G(n,\ell) &= \frac{\ell}{2} - \sum_{i=1}^{\lfloor \ell/2 \rfloor} \alpha(P_i) - \Delta(n,\ell) \\ &= \frac{\ell}{2} - \sum_{i=1}^{a} \alpha(P_i) - \sum_{i=a+1}^{\lfloor \ell/2 \rfloor} \alpha(P_i) - \Delta(n,\ell) \\ &\geq \frac{\ell}{2} - a - (\lfloor \ell/2 \rfloor - a)\alpha(P_{a+1}) - \Delta(n,\ell). \end{aligned} \tag{4.2}$$

The lemma clearly holds if ℓ is even. When ℓ is odd, using the fact that $\Delta(n,\ell) = \alpha(1/2)/2$, we deduce from (4.2) and Lemma 2.3 that

$$\begin{aligned} G(n,\ell) &\geq \frac{\ell}{2} - a - \left(\frac{\ell-1}{2} - a\right)\alpha(P_{a+1}) - \frac{1}{2}\alpha(\frac{1}{2}) \\ &= \left(\frac{\ell}{2} - a\right)(1 - \alpha(P_{a+1})) + \frac{1}{2}\left(\alpha(P_{a+1}) - \alpha(\frac{1}{2})\right) \\ &\geq \left(\frac{\ell}{2} - a\right)(1 - \alpha(P_{a+1})). \end{aligned}$$

The lemma is proved. ■

Next, we need to approximate each p_j and then P_{a+1}.

Lemma 4.2 *Let $j \leq \ell/2$ be a nonnegative integer and let $h = \frac{\ell}{2} - j$. If $4h \leq \ell$, then*

$$p_j \leq \sqrt{\frac{2}{\pi\ell}}\ \exp(\frac{-2h^2}{\ell})\frac{\exp(\frac{1}{12\ell} + \frac{2h^4}{3\ell^3})}{\sqrt{1 - (2h/\ell)^2}},$$

Proof. By Stirling's formula, we have

$$\begin{aligned} p_j = \frac{\ell!}{j!(\ell-j)!2^\ell} &\leq \frac{\sqrt{2\pi\ell}\left(\frac{\ell}{e}\right)^\ell e^{1/(12\ell)}}{\sqrt{2\pi j}\left(\frac{j}{e}\right)^j \sqrt{2\pi(\ell-j)}\left(\frac{\ell-j}{e}\right)^{\ell-j} 2^\ell} \\ &= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{\ell}{j(\ell-j)}}\left(\frac{\ell}{2j}\right)^j\left(\frac{\ell}{2(\ell-j)}\right)^{\ell-j} e^{1/(12\ell)}. \end{aligned} \tag{4.3}$$

First, it is clear that

$$\sqrt{\frac{\ell}{j(\ell-j)}} = \sqrt{\frac{\ell}{\left(\frac{\ell}{2} - h\right)\left(\frac{\ell}{2} + h\right)}} = \frac{2}{\sqrt{\ell}} \cdot \frac{1}{\sqrt{1 - (2h/\ell)^2}}. \tag{4.4}$$

Then, when $4h \le \ell$, we deduce from Lemma 2.4 that

$$\begin{aligned}
\ln\left(\frac{\ell}{2j}\right)^j \left(\frac{\ell}{2(\ell-j)}\right)^{\ell-j} &= -j\ln\left(1-\frac{2h}{\ell}\right) - (\ell-j)\ln\left(1+\frac{2h}{\ell}\right) \\
&\le -j\left(-\frac{2h}{\ell} - \frac{2h^2}{\ell^2} - \frac{8h^3}{3\ell^3} - \frac{8h^4}{\ell^4}\right) - (\ell-j)\left(\frac{2h}{\ell} - \frac{2h^2}{\ell^2} + \frac{8h^3}{3\ell^3} - \frac{4h^4}{\ell^4}\right) \\
&= (-2h)\frac{2h}{\ell} + \ell\frac{2h^2}{\ell^2} + (-2h)\frac{8h^3}{3\ell^3} + (\frac{3}{2}\ell - h)\frac{4h^4}{\ell^4} \qquad (4.5) \\
&= -\frac{2h^2}{\ell} + \frac{2h^4}{3\ell^3} - \frac{4h^5}{\ell^4} \\
&\le -\frac{2h^2}{\ell} + \frac{2h^4}{3\ell^3}
\end{aligned}$$

and thus the lemma follows from (4.3), (4.4), and (4.5). ∎

Lemma 4.3 *Suppose* $\frac{\ell}{2} - z\sqrt{\ell} \le j \le \frac{\ell}{2}$ *and* $\ell \ge 16z^2$. *Then*

$$p_j \le (1+\beta)\sqrt{\frac{2}{\pi\ell}}\, e^{-(\ell-2j)^2/(2\ell)}$$

where

$$\beta = \frac{\exp(\frac{8z^4+1}{12\ell})}{\sqrt{1-\frac{4z^2}{\ell}}} - 1$$

Proof. Let $h = \frac{\ell}{2} - j$. Then we have $4h \le 4z\sqrt{\ell} = \sqrt{16z^2\ell} \le \ell$ and so Lemma 4.2 applies. Notice that $-2h^2/\ell = -(\ell-2j)^2/(2\ell)$, and

$$\frac{\exp(\frac{1}{12\ell} + \frac{2h^4}{3\ell^3})}{\sqrt{1-(2h/\ell)^2}} \le \frac{\exp(\frac{1}{12\ell} + \frac{2(z\sqrt{\ell})^4}{3\ell^3})}{\sqrt{1-(2z\sqrt{\ell}/\ell)^2}} = \beta + 1.$$

Thus the result follows from Lemma 4.2. ∎

Lemma 4.4 *If* $\ell \ge 16z^2$, *then* $P_{a+1} \ge (1+\beta)\left(\Phi(-2z) - \frac{2\sqrt{\ell}+3}{\sqrt{2\pi\ell^3}}\right) - \frac{\beta}{2}$.

Proof. Let

$$\phi(x) = \frac{1}{\sqrt{2\pi\ell}} e^{-x^2/(2\ell)}$$

and let $b = \lceil \ell/2 \rceil - a - 1$. We first consider the case when ℓ is even. Since

$$\sum_{j=0}^{\ell} p_j = 1,$$

and $p_j = p_{\ell-j}$, for every $j = 0, 1, ..., (\ell-2)/2$, it follows that

$$\sum_{j=0}^{(\ell-2)/2} p_j + \frac{1}{2}p_{\ell/2} = \frac{1}{2}.$$

By (3.2) and Lemma 4.3, we have

$$
\begin{aligned}
P_{a+1} = \sum_{j=0}^{a} p_j &= \frac{1}{2} - \frac{1}{2}p_{\ell/2} - \sum_{j=a+1}^{(\ell-2)/2} p_j \\
&\geq \frac{1}{2} - (1+\beta)\phi(0) - 2(1+\beta)\sum_{j=a+1}^{(\ell-2)/2} \phi(\ell - 2j) \\
&= (1+\beta)\left(\frac{1}{2} - \phi(0) - 2\sum_{j=1}^{b}\phi(2j)\right) - \frac{\beta}{2}.
\end{aligned} \tag{4.6}
$$

Observe that, by Lemma 2.2 and (4.1),

$$
\begin{aligned}
\frac{1}{2} - \phi(0) - 2\sum_{j=1}^{b}\phi(2j) &= \Phi(-(2b+1)/\sqrt{\ell}) + \int_0^{2b+1}\phi(x)dx - \phi(0) - 2\sum_{j=1}^{b}\phi(2j) \\
&\geq \Phi(-2z) + \int_0^{2b+1}\phi(x)dx - \phi(0) - 2\sum_{j=1}^{b}\phi(2j).
\end{aligned} \tag{4.7}
$$

Let $2k$ be the largest even integer not exceeding $\sqrt{\ell}$. Then

$$
\begin{aligned}
A &:= \int_0^{2b+1}\phi(x)dx - \phi(0) - 2\sum_{j=1}^{b}\phi(2j) \\
&= \sum_{j=0}^{k-1}\left(\int_{2j}^{2j+1}(\phi(x) - \phi(2j))dx + \int_{2j+1}^{2j+2}(\phi(x) - \phi(2j+2))dx\right) \\
&\quad + \int_{2k}^{2k+1}(\phi(x) - \phi(2k))dx + \int_{2k+1}^{2k+2}(\phi(x) - \phi(2k+2))dx \\
&\quad + \int_{2k+2}^{2k+3}(\phi(x) - \phi(2k+2))dx + \sum_{j=k+1}^{b-1}\int_{2j+1}^{2j+3}(\phi(x) - \phi(2j+2))dx.
\end{aligned} \tag{4.8}
$$

It is straightforward to verify that $\phi''(x) < 0$ over $(0, \sqrt{\ell})$, and $\phi''(x) > 0$ over $(\sqrt{\ell}, \infty)$. Thus we conclude from Lemma 2.6 and the fact that $\phi(x)$ is decreasing on $(0, \infty)$ that

$$
\begin{aligned}
A &\geq 0 + \int_{2k}^{2k+1}(\phi(2k+1) - \phi(2k))dx + \int_{2k+1}^{2k+2}(\phi(2k+2) - \phi(2k+2))dx \\
&\quad + \int_{2k+2}^{2k+3}(\phi(2k+3) - \phi(2k+2))dx + 0 \\
&= (\phi(2k+1) - \phi(2k)) + (\phi(2k+3) - \phi(2k+2)) \\
&= \frac{1}{\sqrt{2\pi\ell}}\left((e^{-\frac{4k^2}{2\ell} - \frac{4k+1}{2\ell}} - e^{-\frac{4k^2}{2\ell}}) + (e^{-\frac{(2k+2)^2}{2\ell} - \frac{4k+5}{2\ell}} - e^{-\frac{(2k+2)^2}{2\ell}})\right).
\end{aligned} \tag{4.9}
$$

By Lemma 2.5, it follows that

$$
A \geq \frac{1}{\sqrt{2\pi\ell}}(-\frac{4k+1}{2\ell} - \frac{4k+5}{2\ell}) = -\frac{4k+3}{\sqrt{2\pi\ell^3}} \geq -\frac{2\sqrt{\ell}+3}{\sqrt{2\pi\ell^3}},
$$

and thus the lemma follows from (4.6), (4.7), (4.8) and (4.9).

The situation for odd ℓ is similar. First,

$$\begin{aligned}
P_{a+1} = \sum_{j=0}^{a} p_j &= \frac{1}{2} - \sum_{j=a+1}^{(\ell-1)/2} p_j \\
&\geq \frac{1}{2} - 2(1+\beta) \sum_{j=a+1}^{(\ell-1)/2} \phi(\ell - 2j) \\
&= \frac{1}{2} - 2(1+\beta) \sum_{j=1}^{b} \phi(2j-1) \\
&= (1+\beta)\left(\frac{1}{2} - 2\sum_{j=1}^{b} \phi(2j-1)\right) - \frac{\beta}{2} \\
&\geq (1+\beta)\left(\Phi(-2z) + \int_0^{2b} \phi(x)dx - 2\sum_{j=1}^{b} \phi(2j-1)\right) - \frac{\beta}{2}.
\end{aligned} \tag{4.10}$$

Let $2k+1$ be the largest odd integer not exceeding $\sqrt{\ell}$. Then

$$\begin{aligned}
B &:= \int_0^{2b} \phi(x)dx - 2\sum_{j=1}^{b} \phi(2j-1) \\
&= \int_0^{1} (\phi(x) - \phi(1))dx + \sum_{j=0}^{k-1} \left(\int_{2j+1}^{2j+2} (\phi(x) - \phi(2j+1))dx + \int_{2j+2}^{2j+3} (\phi(x) - \phi(2j+3))dx \right) \\
&\quad + \int_{2k+1}^{2k+2} (\phi(x) - \phi(2k+1))dx + \int_{2k+2}^{2k+3} (\phi(x) - \phi(2k+3))dx \\
&\quad\quad + \int_{2k+3}^{2k+4} (\phi(x) - \phi(2k+3))dx + \sum_{j=k+2}^{b-1} \int_{2j}^{2i+2} (\phi(x) - \phi(2j+1))dx \\
&\geq (\phi(2k+2) - \phi(2k+1)) + (\phi(2k+4) - \phi(2k+3)) \\
&\geq -\frac{4k+5}{\sqrt{2\pi\ell^3}} \\
&\geq -\frac{2\sqrt{\ell}+3}{\sqrt{2\pi\ell^3}}.
\end{aligned} \tag{4.11}$$

Therefore, by (4.10) and (4.11), the lemma also holds for odd t. The proof is complete. ∎

Lemma 4.5 *Suppose* $\ell \geq \max\{5,\ (z^4+2)/\Phi(-2z)\}$. *Then* $P_{a+1} \geq \Phi(-2z) - \dfrac{z^4+2}{\ell}$.

Proof. By Lemma 2.7, we have $\ell \geq 16z^2$. Also observe that

$$\Phi(-2z) \geq \frac{z^4+2}{\ell} \geq \frac{2}{\ell} \geq \frac{5}{\sqrt{2\pi\ell^2}} \geq \frac{2+3\ell^{-1/2}}{\sqrt{2\pi\ell^2}} = \frac{2\sqrt{\ell}+3}{\sqrt{2\pi\ell^3}}.$$

Thus the lower bound in Lemma 4.4 can be simplified as

$$P_{a+1} \geq \Phi(-2z) - \frac{2\sqrt{\ell}+3}{\sqrt{2\pi\ell^3}} - \frac{\beta}{2}. \tag{4.12}$$

By taking $\lambda = (8z^4+1)/12$, $\mu = 4z^2$, and $x = 1/\ell$, we deduce from Lemma 2.8 and Lemma 2.7 that

$$\beta \le 1.16 \left(\frac{8z^4+1}{12} + 2z^2 \right) \frac{1}{\ell}.$$

Therefore,

$$\frac{2\sqrt{\ell}+3}{\sqrt{2\pi\ell^3}} + \frac{\beta}{2} \le \frac{1}{\ell} \left(\frac{1.16}{3} z^4 + 1.16z^2 + \frac{2+\frac{3}{\sqrt{5}}}{\sqrt{2\pi}} + \frac{1.16}{12} \right) \le \frac{z^4+2}{\ell},$$

and thus the lemma follows from (4.12). ∎

Proof of Theorem 1.1. We first prove that the assumption $\ell \ge \ell(n)$ implies $\ell \ge \max\{5,\ (z^4+2)/\Phi(-2z)\}$, if we take

$$z = \sqrt{\frac{(1-\epsilon)\ln n}{2}}.$$

Notice that $\ell(2) \ge \sqrt{\pi/8}\ ((\ln 2)^2+8)\ (\sqrt{2\ln 2}+1) \approx 11.6$ and thus $\ell \ge 5$ holds. Furthermore, by Lemma 2.2, we have

$$\begin{aligned} \frac{z^4+2}{\Phi(-2z)} &\le (z^4+2)\sqrt{2\pi}\ (2z+1)\ e^{2z^2} \\ &= \left(\left(\frac{(1-\epsilon)\ln n}{2} \right)^2 + 2 \right) \sqrt{2\pi}\ (\sqrt{2(1-\epsilon)\ln n}+1)\ n^{1-\epsilon} \\ &\le \ell(n). \end{aligned}$$

Therefore, the assumptions in Lemma 4.5 are satisfied. Consequently, by Lemma 2.2 again, we have

$$\begin{aligned} P_{a+1} &\ge \Phi(-2z) - \frac{z^4+2}{\ell} \\ &\ge \frac{e^{-2z^2}}{\sqrt{2\pi}\ (2z+1)} - \frac{z^4+2}{\ell} \\ &= \frac{1}{\sqrt{2\pi}\ (\sqrt{2(1-\epsilon)\ln n}+1)\ n^{1-\epsilon}} - \frac{\left(\frac{(1-\epsilon)\ln n}{2}\right)^2+2}{\ell} \\ &\ge \delta. \end{aligned} \tag{4.13}$$

In addition, from (1.4) it is clear that $\delta \ge 0$. Thus we deduce from Lemma 4.1 and Lemma 2.3 that

$$G(n,\ell) \ge (\frac{\ell}{2} - a)(1-\alpha(\delta)) \ge (z\sqrt{\ell} - \frac{1}{2})(1-\alpha(\delta)).$$

The proof is complete. ∎

Acknowledgment.

The authors are grateful to their friend Steve Seiden, who introduced this problem to them. They wish to acknowledge that they have benefited from discussions with Steve. The authors would also like to thank the two anonymous referees for their valuable suggestions. This Research is partially supported by NSF grant DMS-9970329 and AFOSR grant F49620-01-1-0264.

References

[1] N. Cesa-Bianchi, Y. Freund, D. P. Helmbold, and M. Warmuth, On-line Prediction and Conversion Strategiese, *Machine Learning*, **25** (1996), 717-110.

[2] N. Cesa-Bianchi, Y. Freund, D. Haussler, D. P. Helmbold, and R. E. Schapire, How to use expert advice, *Journal of the ACM*, Vol. 44, No. 3, May 1997, pp. 427-485.