Inf. Process. Lett. 187 (2025) 106516

Contents lists available at ScienceDirect

Information Processing Letters

journal homepage: www.elsevier.com/locate/ipl

Leveraging parameterized Chernoff bounds for simplified algorithm analyses

Michael Dillencourt[a], Michael T. Goodrich[a,*], Michael Mitzenmacher[b]

[a] Univ. of California, Irvine, CA, USA
[b] Harvard Univ., Cambridge, MA, USA

ABSTRACT

In this paper, we derive parameterized Chernoff bounds and show their applications for simplifying the analysis of some well-known probabilistic algorithms and data structures. The parameterized Chernoff bounds we provide give probability bounds that are powers of two, with a clean formulation of the relation between the constant in the exponent and the relative distance from the mean. In addition, we provide new simplified analyses with these bounds for hash tables, randomized routing, and a simplified, non-recursive adaptation of the Floyd-Rivest selection algorithm.

1. Introduction

Chernoff bounds [4,15] have been shown to be useful for analyzing a wide variety of different probabilistic algorithms and data structures. Examples of their use can be found in textbooks by Alon and Spencer [1], Motwani and Raghavan [19], and Mitzenmacher and Upfal [18]. Some of the most well-known Chernoff bounds provide a bound on the probability that a sum, $X = \sum_{i=1}^{n} X_i$, of independent random variables taking values in $\{0,1\}$ has a value sufficiently far away from its expected value $\mu = E[X]$. The multiplicative form of a Chernoff bound for such a random variable X is commonly stated as follows.

Theorem 1 *(See [1,14,18,19]). For any* $\delta > 0$,

$$\Pr(X > (1+\delta)\mu) < \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}.$$

Also, for any $0 < \delta < 1$,

$$\Pr(X < (1-\delta)\mu) < \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu}.$$

These formulas are often unwieldy to use in practice, however. Hence, algorithmic analyses often use simpler Chernoff bounds, with the following being common.

Theorem 2 *(See [1,2,14,18,19,22]).*

$$\Pr(X > (1+\delta)\mu) < e^{-\delta^2\mu/(2+\delta)}, \quad \textit{for } \delta > 0, \tag{1}$$

$$\Pr(X < (1-\delta)\mu) < e^{-\delta^2\mu/2}, \quad \textit{for } 0 < \delta < 1. \tag{2}$$

As one example demonstrating how influential these bounds have been, we note that a paper in *Information Processing Letters* (IPL) by Hagerup and Rüb [14], which includes simple bounds like those in Theorem 2, has been cited hundreds of times. Nevertheless, as simple as the above "simplified" Chernoff bounds are, they have the following drawbacks with respect to algorithmic applications:

1. The probabilities in Theorem 2 are powers of Euler's number, $e \approx 2.71828183\ldots$, rather than of 2. In algorithmic applications, however, it is often preferred to express probabilities as powers of 2. Indeed, some algorithmic researchers will apply a simplified Chernoff bound, as in Theorem 2, and then convert the resulting probability to a power of two using the crude inequality, $2 \le e$, which results in a loss of accuracy; e.g., see Elsässer and Sauerwald [10].
2. The probabilities in Theorem 2 involve annoying δ^2 terms, due in part to the need for the bounds to hold for values of δ very close to 0, whereas algorithmic analyses are generally indifferent to very small values of δ. Indeed, Chernoff bounds are designed as upper bounds for the ***tails*** of random variables.

In this paper, we build on a recent IPL paper by Dillencourt and Goodrich [8], to derive simple ***parameterized*** Chernoff bounds, with the following goals:

[*] Corresponding author.
E-mail addresses: mbdillen@uci.edu (M. Dillencourt), goodrich@uci.edu (M.T. Goodrich), michaelm@eecs.harvard.edu (M. Mitzenmacher).

https://doi.org/10.1016/j.ipl.2024.106516
Received 10 November 2023; Received in revised form 4 June 2024; Accepted 7 June 2024

Available online 12 June 2024

1. Characterize probabilities as powers of 2, with a variable parameter, x, in the exponent.[1]
2. Avoid δ^2 terms where possible, for example by loosening the requirement that the bounds hold for values of δ very close to 0.

1.1. Related prior work

In terms of prior work, there is a notable upper-tail Chernoff bound where the error bound is a power of two, from a book by Mitzenmacher and Upfal [18] and the paper by Hagerup and Rüb [14]:

Theorem 3 *(Mitzenmacher and Upfal [18] (p. 69) and Hagerup and Rüb [14]).*

$$\Pr(X > R) < 2^{-R}, \quad \text{for } R \geq 6\mu.$$

In addition, Motwani and Raghavan [19] leave as an exercise to prove $\Pr(X > R) < 2^{-R}$, for $R \geq 2e\mu$, which is a slightly better condition, since $2e \approx 5.43656$. Although this Chernoff bound is useful, and it partially satisfies the two goals given above for simplified Chernoff bounds, it does not always result in the best bounds. It can be improved with parameterized Chernoff bounds, as we show.

In more recent work, Shiu [22] derives tighter Chernoff bounds in the fashion of Theorem 2, but these bounds are not parameterized and do not satisfy either of the two goals outlined above for algorithmic applications. In addition, Dillencourt and Goodrich [8] derive some simplified Chernoff bounds that partially satisfy the two goals outlined above for algorithmic applications, but their bounds are not parameterized; hence, they are not applicable for the algorithm analysis applications we address in this paper.

1.2. Our results

In this paper, we derive parameterized Chernoff bounds with probability bounds that are powers of two and that, with one exception, avoid δ^2 terms, for reasonable values of $\delta > 0$, and are parameterized with a single parameter, $x > 0$. We also provide algorithmic applications of our parameterized Chernoff bounds, but we believe these are just the tip of the iceberg in terms of simplified analyses that are possible. We stress that the main contributions of this paper are for ***parameterized*** Chernoff bounds, and that none of our Chernoff bounds are tighter than those given in Theorem 1. Instead, we argue and show by example that the type of parameterized Chernoff bounds provided in this paper are easier to use for algorithmic applications than the Chernoff bounds of Theorem 1. Further, the Chernoff bounds provided in this paper are often tighter than the bounds of Theorems 2 and 3.

In addition, we show how our parameterized Chernoff bounds can be used to provide new simplified analyses of hash tables. Also, we provide a new simplified, non-recursive adaptation of the Floyd-Rivest selection algorithm, and show how to use our parameterized Chernoff bounds to show that this randomized method for finding the kth smallest of n comparable items uses $n + \min\{k, n-k\} + o(n)$ comparisons with high probability.

2. The Lambert W function

Since we will derive parameterized Chernoff bounds by making use of the Lambert W function, we first review this function. The Lambert W function is defined by the rule that $W(z) = w$ if and only if w satisfies the following equation:

$$we^w = z,$$

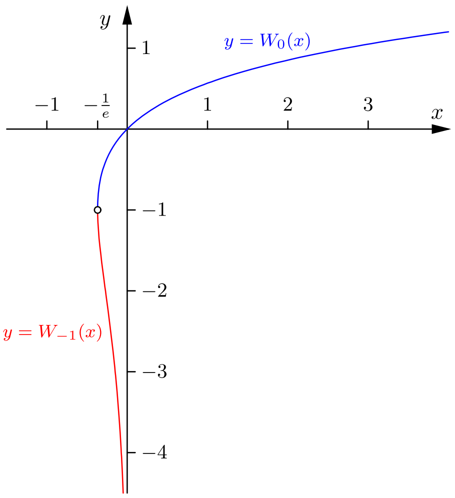

Fig. 1. The two real branches of the Lambert W function. Image Copyright © 2022 Michael Dillencourt; used with permission.

where z is a complex number. See, for example, Corless, Gonnet, Hare, Jeffrey, and Knuth [5], Barry, Culligan-Hensley, and Barry [3], Corless, Jeffrey, and Knuth [6], or Iacono and Boyd [16].

Technically, W is not a function. Hence its real-valued expression is partitioned into two branches, for $x \geq -1/e$: $W_0(x)$, which is called the ***principal branch*** and is always greater than or equal to -1, and $W_{-1}(x)$, which is called the ***non-principal branch*** and is always less than or equal to -1. A plot of the two real branches is shown in Fig. 1. The two branches split at $(-\frac{1}{e}, -1)$. Thus, $W_0(ye^y) = y$ for $y \geq -1$, and $W_{-1}(ye^y) = y$ for $y \leq -1$.

The Lambert W function has several applications in algorithm analysis [5]. For example, if we define the function, $\lambda(x)$, to be

$$\lambda(x) = \frac{-1}{W_0\left(\frac{-1}{2^x e}\right)},$$

then we can interpret the work of Devroye [7] and Reed [20] as showing that the expected height of a randomly-constructed binary search tree is $\lambda(1)\log_2 n + O(1) \approx 4.31107 \log_2 n$.

The Lambert W function cannot be expressed in terms of elementary functions [5]; hence, evaluating it typically requires the use of a numerical algorithm [16]. For example, its principal branch, W_0, has the following Taylor series expansion around 0 (see, e.g., Corless, Gonnet, Hare, Jeffrey, and Knuth [5]), for $-1/e < x < \infty$:

$$W_0(x) = \sum_{i=1}^{\infty} \frac{(-i)^{i-1}}{i!} x^i = x - x^2 + \tfrac{3}{2}x^3 - \tfrac{8}{3}x^4 + \tfrac{125}{24}x^5 - \cdots.$$

3. Parameterized Chernoff bounds

In this section, we derive some Chernoff bounds with probabilities that are powers of two and that are expressed using a parameter, x. Interestingly, our parameterized Chernoff bounds provide another surprising application of the $\lambda(x)$ function defined above.

Theorem 4. *Let $X_1, X_2, \ldots, X_n$ be independent random variables taking values in $\{0,1\}$. Let $X = \sum_{i=1}^n X_i$ and let $\mu = E[X]$ denote X's expected value. Then*

$$\Pr(X > R) < 2^{-xR},$$

for $x > 0$ and $R \geq \lambda(x)\mu$, where $\lambda(x) = -1/W_0(-1/(2^x e))$ and $W_0(z)$ is the principal branch of the Lambert W function.

Proof. From the general form of the Chernoff bound of Theorem 1, taking $R = (1+\delta)\mu$,

[1] In some cases, we may also derive bounds for specific values of x.

$$\Pr(X > R) = \Pr(X > (1+\delta)\mu) < \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}.$$

In order for this probability to be at most 2^{-xR}, for $x > 0$, we need $1+\delta \geq 2^x e^{\delta/(1+\delta)}$. Setting $z = 1+\delta$, the breakpoint for this inequality occurs for z satisfying

$$z = 2^x e^{\frac{z-1}{z}},$$

which can be rewritten as

$$ze^{\frac{1}{z}} = 2^x e,$$

or

$$-\left(\frac{1}{z}\right)e^{-\frac{1}{z}} = \frac{-1}{2^x e}.$$

Thus, we have an equation in the form of the Lambert W function. To ensure $\delta > 0$, we must have $z > 1$ and hence $-1/z > -1$. So we choose the principal branch of the Lambert W function to obtain $(-1/z) = W_0(-1/(2^x e))$, or $z = -1/W_0(-1/(2^x e))$, that is, $z = \lambda(x)$. □

If one desires a Chernoff bound expressed in terms of factors x and $(1+\delta)$ times μ, Theorem 4 can be restated as follows.

Theorem 5. *Let $X_1, X_2, \ldots, X_n$ be independent random variables taking values in $\{0,1\}$. Let $X = \sum_{i=1}^n X_i$ and let $\mu = E[X]$ denote X's expected value. Then*

$$\Pr(X > (1+\delta)\mu) < 2^{-x(1+\delta)\mu},$$

for $x > 0$ and $\delta \geq \lambda(x) - 1$, where $\lambda(x) = -1/W_0(-1/(2^x e))$ and $W_0(z)$ is the principal branch of the Lambert W function.

Proof. Let $R = (1+\delta)\mu$ and apply Theorem 4. □

Since the Lambert W function cannot be expressed in terms of elementary functions, most uses of Theorems 4 and 5 will likely be for specific values of x and $\lambda(x)$. The following theorem contains some examples.

Theorem 6. *Let $X_1, X_2, \ldots, X_n$ be independent random variables taking values in $\{0,1\}$. Let $X = \sum_{i=1}^n X_i$ and let $\mu = E[X]$ denote X's expected value. Then*

$$\Pr(X > (1+\delta)\mu) < 2^{-(1+\delta)\mu/256}, \quad \text{for } \delta \geq 0.07735. \tag{3}$$

$$\Pr(X > (1+\delta)\mu) < 2^{-(1+\delta)\mu/128}, \quad \text{for } \delta \geq 0.11172. \tag{4}$$

$$\Pr(X > (1+\delta)\mu) < 2^{-(1+\delta)\mu/64}, \quad \text{for } \delta \geq 0.16285. \tag{5}$$

$$\Pr(X > (1+\delta)\mu) < 2^{-(1+\delta)\mu/32}, \quad \text{for } \delta \geq 0.24063. \tag{6}$$

$$\Pr(X > (1+\delta)\mu) < 2^{-(1+\delta)\mu/16}, \quad \text{for } \delta \geq 0.36278. \tag{7}$$

$$\Pr(X > (1+\delta)\mu) < 2^{-(1+\delta)\mu/8}, \quad \text{for } \delta \geq 0.56405. \tag{8}$$

$$\Pr(X > (1+\delta)\mu) < 2^{-(1+\delta)\mu/6}, \quad \text{for } \delta \geq 0.68619. \tag{9}$$

$$\Pr(X > (1+\delta)\mu) < 2^{-(1+\delta)\mu/4}, \quad \text{for } \delta \geq 0.92051. \tag{10}$$

$$\Pr(X > (1+\delta)\mu) < 2^{-(1+\delta)\mu/3}, \quad \text{for } \delta \geq 1.15187. \tag{11}$$

$$\Pr(X > (1+\delta)\mu) < 2^{-(1+\delta)\mu/2}, \quad \text{for } \delta \geq 1.62729. \tag{12}$$

$$\Pr(X > (1+\delta)\mu) < 2^{-(1+\delta)\mu}, \quad \text{for } \delta \geq 3.31107. \tag{13}$$

$$\Pr(X > (1+\delta)\mu) < 2^{-2(1+\delta)\mu}, \quad \text{for } \delta \geq 8.82044. \tag{14}$$

$$\Pr(X > (1+\delta)\mu) < 2^{-3(1+\delta)\mu}, \quad \text{for } \delta \geq 19.72173. \tag{15}$$

$$\Pr(X > (1+\delta)\mu) < 2^{-4(1+\delta)\mu}, \quad \text{for } \delta \geq 41.48069. \tag{16}$$

In addition, we have the following parameterized Chernoff bound, which provides an expression for its conditions in terms of elementary functions of the parameter, x, and is also an improved bound over Theorem 3 for $x = 1$.

Theorem 7. *Let $X_1, X_2, \ldots, X_n$ be independent random variables taking values in $\{0,1\}$. Let $X = \sum_{i=1}^n X_i$ and let $\mu = E[X]$ denote X's expected value. Then*

$$\Pr(X > R) < 2^{-xR}, \quad \text{for } x > 0 \text{ and } R \geq (2^x e - 1)\mu.$$

Proof. The proof follows from Theorem 4, provided we can show that, for $x > 0$,

$$2^x e - 1 \geq \lambda(x) = -1/W_0(-1/(2^x e)).$$

For convenience, let $z = -1/(2^x e)$, and note that $z \in (-1/e, 0)$. We then want to show that

$$-1/W_0(z) \leq -(1/z) - 1,$$

or equivalently

$$W_0(z) - z/(z+1) \leq 0.$$

We examine the Taylor expansions, around 0, recalling (see, e.g., [6,16,23])

$$W_0(z) = \sum_{i=1}^{\infty} \frac{(-i)^{i-1}}{i!} z^i,$$

and

$$z/(z+1) = \sum_{i=1}^{\infty} (-1)^{i-1} z^i,$$

for $z \in (-1/e, 0)$. It follows that

$$W_0(z) - z/(z+1) = \sum_{i=1}^{\infty} \frac{(-i)^{i-1} - (-1)^{i-1}(i!)}{i!} z^i.$$

We see that the coefficient of z^i is 0 for $i = 1, 2$. For larger i, the coefficient is positive for odd i and negative for even i, since $i^{i-1} > i!$, for $i \geq 3$. As z is negative, each term is negative; hence, we have $W_0(z) \leq z/(z+1)$, as desired. □

Suppose instead of being mutually independent, the n 0-1 random variables in a sum, $X = \sum_{i=1}^n X_i$, are only k-wise independent. We can derive a parameterized Chernoff bound in this case, provided k is large enough:

Theorem 8. *Let $X_1, X_2, \ldots, X_n$ be k-wise independent random variables taking values in $\{0,1\}$, for $k \geq \lceil \delta\mu \rceil$, for $\delta > 0$. Let $X = \sum_{i=1}^n X_i$ and let $\mu = E[X]$ denote X's expected value. Then the bounds of Theorems 4, 5, 6, and 7 hold for X.*

Proof. Schmidt, Siegel, and Srinivasan [21] show that the upper-tail bounds of Theorem 1 hold for k-wise independent random variables if $k \geq \lceil \delta\mu \rceil$, for $\delta > 0$. The bounds of Theorems 4, 5, 6, and 7 depend only on the formulation of the upper-tail bound in Theorem 1. □

4. Additional Chernoff bounds for specific parameter values

The main results in this paper are for parameterized Chernoff bounds in terms of a variable, x, but let us also provide some additional Chernoff bounds for specific parameter values.

For example, we have focused primarily on Chernoff bounds for values of δ that are not too close to 0, but there are occasions in which such Chernoff bounds are desired. For such occasions, one can use the following simplified Chernoff bounds, albeit with δ^2 terms.

Theorem 9. *Let $X_1, X_2, \ldots, X_n$ be independent random variables taking values in $\{0,1\}$. Let $X = \sum_{i=1}^n X_i$ and let $\mu = E[X]$ denote X's expected value. Then, for $0 < \delta < 1$ and $\gamma = \log_2 e$,*

$$\Pr(X > (1+\delta)\mu) < 2^{-1.5\gamma\delta^2\mu/(3+\delta)} < 2^{-2\delta^2\mu/(3+\delta)} \quad \textit{and} \tag{17}$$

$$\Pr(X < (1-\delta)\mu) < 2^{-9\gamma\delta^2\mu/(18-6\delta-\delta^2)} < 2^{-2\delta^2\mu/(3-\delta-\delta^2/6)}. \tag{18}$$

Proof. These bounds follow from tightened Chernoff bounds of Shiu [22]. □

Alternatively, in case one would like to find a specific bound for larger values of δ based on a desired simple probability bound using powers of two, Dillencourt and Goodrich [8] establish the following.

Theorem 10 *(Dillencourt and Goodrich [8]). Let $X_1, X_2, \ldots, X_n$ be independent random variables taking values in $\{0,1\}$. Let $X = \sum_{i=1}^n X_i$ and let $\mu = E[X]$ denote X's expected value. Then*

$$\Pr(X > (1+\delta)\mu) < 2^{-\alpha\mu}, \tag{19}$$

holds for fixed $\alpha > 0$ when

$$\delta \ge e^{W_0\left(\frac{\alpha \ln 2 - 1}{e}\right)+1} - 1. \tag{20}$$

Further, if $\delta < 1$, then

$$\Pr(X < (1-\delta)\mu) < 2^{-\beta\mu}, \tag{21}$$

holds for fixed $\beta > 0$ when

$$\delta \ge 1 - e^{W_{-1}\left(\frac{\beta \ln 2 - 1}{e}\right)+1}. \tag{22}$$

We can use Theorem 10 to derive parameterized Chernoff lower-tail bounds for specific constant parameters, as exemplified in the following theorem.

Theorem 11. *Let $X_1, X_2, \ldots, X_n$ be independent random variables taking values in $\{0,1\}$. Let $X = \sum_{i=1}^n X_i$ and let $\mu = E[X]$ denote X's expected value. Then*

$$\Pr(X < (1-\delta)\mu) < 2^{-\mu/2}, \quad \text{for } 1 > \delta \ge 0.7064. \tag{23}$$

$$\Pr(X < (1-\delta)\mu) < 2^{-\mu/3}, \quad \text{for } 1 > \delta \ge 0.5974. \tag{24}$$

$$\Pr(X < (1-\delta)\mu) < 2^{-\mu/4}, \quad \text{for } 1 > \delta \ge 0.5276. \tag{25}$$

$$\Pr(X < (1-\delta)\mu) < 2^{-\mu/8}, \quad \text{for } 1 > \delta \ge 0.3863. \tag{26}$$

$$\Pr(X < (1-\delta)\mu) < 2^{-\mu/16}, \quad \text{for } 1 > \delta \ge 0.2796. \tag{27}$$

$$\Pr(X < (1-\delta)\mu) < 2^{-\mu/32}, \quad \text{for } 1 > \delta \ge 0.2008. \tag{28}$$

5. Revisiting balls-in-bins analyses for hash tables

In this section, we show some applications of parameterized Chernoff bounds to balls-in-bins problems for hash-table applications. We start with a simple proof of a well-known result, which has a textbook proof that uses Stirling's approximation rather than a Chernoff bound and is arguably less simple, e.g., see Mitzenmacher and Upfal [18, p. 100].

Theorem 12. *If n balls are thrown independently and uniformly at random into n bins, then the probability that the largest bin has more than $3\log n/\log\log n$ balls is at most $1/n$ for sufficiently large n.*

Proof. Let $X = \sum_{i=1}^n X_i$ be a random variable for the number of balls thrown into bin 1, where X_i is a 0-1 random variable that is 1 if and only if ball i is thrown into bin 1. Thus, $E[X_i] = 1/n$; hence, $\mu = E[X] = 1$. Taking $x = \log\log n - \log\log\log n$ and $R = 2^x 3\mu > 2^x e\mu$, we can apply Theorem 7 as follows:

$$\begin{aligned}\Pr(X > 3\log n/\log\log n) &= \Pr(X > R)\\ &< 2^{-xR}\\ &= 2^{-(\log\log n - \log\log\log n)3\log n/\log\log n}\\ &\le 2^{-2\log n}\\ &= \frac{1}{n^2},\end{aligned}$$

for suitably large n. The proof follows, then, by a union bound. □

In fact, we can prove something even stronger.

Theorem 13. *If n balls are thrown independently and uniformly at random into $n/\log\log n$ bins, then the probability that the largest bin has more than $3\log n/\log\log n$ balls is at most $1/(n\log\log n)$ for sufficiently large n.*

Proof. The proof is similar to that for Theorem 12. Let $X = \sum_{i=1}^n X_i$ be a random variable for the number of balls thrown into bin 1, where X_i is a 0-1 random variable that is 1 if and only if ball i is thrown into bin 1. Thus, $E[X_i] = (\log\log n)/n$; hence, $\mu = E[X] = \log\log n$. Taking $x = \log\log n - 2\log\log\log n$ and $R = 2^x 3\mu > 2^x e\mu$, we can apply Theorem 7 as follows:

$$\begin{aligned}\Pr(X > 3\log n/\log\log n) &= \Pr(X > R)\\ &< 2^{-xR}\\ &= 2^{-(\log\log n - 2\log\log\log n)3\log n/\log\log n}\\ &\le 2^{-2\log n}\\ &= \frac{1}{n^2},\end{aligned}$$

for suitably large n. The proof follows, then, by a union bound. □

We can also easily prove other interesting balls-in-bins results, with negligible probabilities,[2] such as the following.

Theorem 14. *If n balls are thrown independently and uniformly at random into n bins, then the probability that the largest bin has more than $3\log n$ balls is negligible.*

Proof. The proof is similar to that for Theorem 12. Let $X = \sum_{i=1}^n X_i$ be a random variable for the number of balls thrown into bin 1, where X_i is a 0-1 random variable that is 1 if and only if ball i is thrown into bin 1. Thus, $E[X_i] = 1/n$; hence, $\mu = E[X] = 1$. Taking $x = \log\log n$ and $R = 2^x 3\mu > 2^x e\mu$, we can apply Theorem 7 as follows:

$$\begin{aligned}\Pr(X > 3\log n) &= \Pr(X > R)\\ &< 2^{-xR}\\ &= 2^{-(\log\log n)3\log n}\\ &= \frac{1}{n^{3\log\log n}}.\end{aligned}$$

The proof follows, then, by a union bound. □

Note that these theorems do not follow from a direct application of Theorem 3.

[2] Recall that a function is ***negligible*** if it tends towards 0 faster than the reciprocal of any polynomial.

6. Some algorithmic applications

In this section, we highlight some improved analyses that are implied by the above bounds.

6.1. Randomized routing in a hypercube

A well-known "textbook" application of Chernoff bounds is for permutation routing in a hypercube; see, e.g., [18,19]. In this problem, every node in a hypercube with $N = 2^n$ nodes starts with a packet, which is sent to another node in the hypercube. We assume each destination node also receives one packet, so we are routing a permutation. At most one packet can cross any edge at any time step. Valiant suggested routing the permutation by using a two-phase randomized routing [24,25]: in a first phase, each packet is sent to a randomly chosen destination, and, in a second phase, each packet continues to its final destination. This randomized routing will route the permutation in $O(n)$ total time steps. We consider here the first phase, as the second phase is entirely similar to analyze.

We follow the framework of Mitzenmacher and Upfal [18], and refer there for additional details. Each node is represented by an n-bit vector, $(x_1, x_2, \ldots, x_n)$. The packet is sent using the ***bit-fixing route***; that is, in sending a packet from node $(x_1, x_2, \ldots, x_n)$ to node $(y_1, y_2, \ldots, y_n)$, we consider the x_i in order, and whenever $x_i \neq y_i$, the packet crosses the edge $(y_1, y_2, \ldots, y_{i-1}, x_i, x_{i+1}, \ldots, x_n)$ to $(y_1, y_2, \ldots, y_{i-1}, y_i, x_{i+1}, \ldots, x_n)$. The key goal is to show that for any possible packet path, P, no more than cn distinct packets are ***active*** with sufficiently high probability. Here, active means that the packet reaches a vertex of P and has the potential to cross an edge of P; that is, if P reaches a node v on path P and the next node w of P differs from v on the jth bit, then when the packet reaches v the jth bit of the packet's path cannot have been processed by the bit-fixing algorithm. With this bound, one can argue that the cn packets cross the edges of P at most $c'n$ times for some other constant c' (since each packet leaves the path P with probability at least $1/2$ at each vertex of P), and the result readily follows.

To start, for $k = 1, \ldots, N$, let H_k be a 0-1 random variable where $H_k = 1$ if the packet starting at node k is active and 0 otherwise. The H_k are independent, and we let $H = \sum H_k$. If neighbors v and w on the path differ in the jth bit, there are only 2^{j-1} possible active packets, as an active packet must begin at a vertex that agrees with v on bits j through n of its address. Further, each such packet reaches v with probability $2^{-(j-1)}$, since each such possible packet must choose a random destination that matches v's first $j-1$ address bits. Hence, the expected number of active packets per vertex is 1, and accordingly, $E[H] \leq n$, as the path P has at most n vertices.

In the textbook by Mitzenmacher and Upfal [18], the Chernoff bound

$$\Pr(H \geq 6n \geq 6E[H]) \leq 2^{-6n}$$

is then applied. With some additional work, it is then shown that each phase takes at most $30n$ steps with probability at least $1 - O(1/N)$, so the two phases complete in at most $60n$ steps with probability at least $1 - O(1/N)$. (There does not seem to have been effort to optimize the constant 60.) We can improve this using Theorem 7 with $x = 0.8$, to find that

$$\Pr(H \geq 4n) \leq 2^{-3.2n},$$

which suffices for the rest of the proof given in Mitzenmacher and Upfal [18] (as long as n is sufficiently large, e.g. $n \geq 10$). To continue, let T be the number of times the H packets cross an edge of P, which bounds the total time a packet could take to traverse P. As each packet at a vertex of P may leave P with probability $1/2$, we see that, conditioned on $H \leq 4n$, the probability that packets cross an edge of P more than $16n$ times is bounded by the probability that a fair coin flipped $20n$ times yields fewer than $4n$ heads. (Each time a packet at a vertex of P has to cross an edge, as its route is random, we model whether the bit flips in that dimension as a coin flip: if the coin is heads, the packet leaves the path and is no longer considered, and if the coin is tails, then the packet follows the edge of P. After $4n$ heads, we are out of packets if $H \leq 4n$, so the probability of at least $16n$ edge traversals on P is bounded by the given probability.) Let Z be the number of heads in $20n$ coin flips. By Theorem 11, Equation (24), for $1 > \delta \geq 0.6$, we have that

$$\Pr(X < (1-\delta)\mu) < 2^{-\mu/3}.$$

Accordingly, for $20n$ coin flips, the number of heads Z satisfies

$$\Pr(Z < 4n) < 2^{-10n/3} < 2^{-3.3n}.$$

We therefore have for any path P, when $n \geq 10$,

$$\begin{aligned}\Pr(T > 16n) &\leq \Pr(T > 16n \mid H > 4n)\Pr(H > 4n) \\ &\quad + \Pr(T > 16n \mid H \leq 4n)\Pr(H \leq 4n) \\ &\leq \Pr(H > 4n) + \Pr(T > 16n \mid H \leq 4n) \\ &\leq \Pr(H > 4n) + \Pr(Z < 4n) \\ &\leq 2^{-3.2n} + 2^{-3.3n} \\ &< 2^{-3n}.\end{aligned}$$

There are at most $N^2 = 2^{2n}$ paths P, so we have that the probability any path has a delay of more than $16n$ steps is at most $1/N$. It follows that this analysis, using the power-of-two Chernoff bounds, improves the bound for two-phase randomized rounding to $32n$ steps while arguably simplifying the proof.

6.2. Analysis of a non-recursive adaptation of the Floyd-Rivest selection algorithm

The Floyd-Rivest selection algorithm is a randomized, recursive method for finding the kth smallest of n comparable items using $n + \min\{k, n-k\} + o(n)$ comparisons with high probability. Although the algorithm itself is simple, its published analyses are not; see, e.g., [12,13,17]. For example, while the algorithm is simple enough to present to undergraduates, we are not aware of lecture notes for its analysis; see, e.g., Eppstein [11].

In this subsection, we provide a simplified non-recursive adaptation of the Floyd-Rivest selection algorithm [12]. Our goal here is to provide an analysis that achieves an optimal number of comparisons for the selection problem, plus lower-order terms, which could be presented to undergraduates.

Our non-recursive adaptation of the Floyd-Rivest selection algorithm [12] is inspired by lecture notes of Eppstein [11] and pseudo-code of Kiwiel [17], but is nevertheless different from both of these methods. Given a set, S, of n distinct comparable elements, and $1 \leq k \leq n$, the following algorithm returns the kth smallest element in S, using a sample size parameter, s, and gap-size parameter, g.

```
Select($S, n, k, s, g$):
1. Choose a random subset, $R$, of $S$ of expected size $s$, by choosing
   each element of $S$ independently with probability $s/n$.
2. Sort $R$.
3. If $ks/n - g > 1$, then let $u$ be the element of rank $\lfloor ks/n - g \rfloor$ in $R$.
4. If $ks/n + g < s$, then let $v$ be the element of rank $\lceil ks/n + g \rceil$ in $R$.
5. If $k \leq n/2$, then:
   (a) Partition $S$ into $S'$, the elements less than or equal to $v$, and
       $G$, the elements greater than $v$.
   (b) If $ks/n - g > 1$, then partition $S'$ into $L$, the elements less than
       $u$, and $M$, the elements greater than or equal to $u$. Otherwise,
       let $L = \emptyset$ and $M = S'$.
```

Else:
(a) Partition S into L, the elements less than u, and S', the elements greater than or equal to u.
(b) If $ks/n + g < s$, then partition S' into M, the elements less than or equal to v, and G, the elements greater than v. Otherwise, let $G = \emptyset$ and $M = S'$.

6. If $|L| < k$ and $|L| + |M| \geq k$, then sort M and return the element of rank $k - |L|$ in M. Otherwise, repeat the entire algorithm from scratch.

Theorem 15. *One can choose s and g so that our non-recursive adaptation of the Floyd-Rivest selection algorithm uses $n + \min\{k, n-k\} + o(n)$ comparisons, with high probability.*

Proof. W.l.o.g., let us assume that $k \leq n/2$. Assuming our non-recursive adaptation of the Floyd-Rivest selection algorithm doesn't restart and we do the sorting steps with a worst-case optimal comparison-based sorting algorithm, like heapsort or mergesort, then it is easy to see that the total number of comparisons the algorithm performs is $n + |L| + O(|R| \log |R|) + O(|M| \log |M|)$. Note that if $ks/n - g \leq 1$, then $|L| = 0$, which simplifies the analysis, so let us consider the more general case when $ks/n - g > 1$, that is, $k > (g+1)n/s$.

Let us bound the probability for the bad outcomes that would cause a restart of the algorithm. First, consider the bad outcome where $|L| \geq k$. This would occur if u has rank greater than k in S. That is, if we let X denote the number of elements in S with rank at most k in S that are chosen to be in R, then $|L| \geq k$ iff $X < ks/n - g$. Noting that $\mu = E[X] = ks/n$,

$$\Pr(X < ks/n - g) = \Pr(X < (1-\delta)\mu),$$

where $\delta = g/\mu = gn/ks$ and $k > (g+1)n/s$, so $0 < \delta < 1$. Thus, by Theorem 9,

$$\Pr(X < (1-\delta)\mu) < 2^{-2\delta^2\mu/3} = 2^{-2g^2/3\mu} = 2^{-2g^2n/3ks} \leq 2^{-4g^2/3s}.$$

Let us choose $s = n^{2/3}$ and $g = n^{1/3} \log^{1/2} n$, which implies that

$$\Pr(X < (1-\delta)\mu) < 2^{-(4/3)\log n} < n^{-4/3}.$$

Thus, with high probability, we avoid this bad case.

Next, let us consider the bad outcome where $|L| + |M| < k$, which occurs if the rank of v in S is less than k. That is, if we again let X denote the number of elements in S with rank at most k in S that are chosen to be in R, then $|L| + |M| < k$ iff $X > ks/n + g$. Noting that $\mu = E[X] = ks/n$,

$$\Pr(X > ks/n + g) = \Pr(X < (1+\delta)\mu),$$

where $\delta = g/\mu = gn/ks$ and $k > (g+1)n/s$, so $0 < \delta < 1$. Thus, by Theorem 9,

$$\Pr(X < (1+\delta)\mu) < 2^{-2\delta^2\mu/3} = 2^{-2g^2/3\mu} = 2^{-2g^2n/3ks} \leq 2^{-4g^2/3s};$$

hence, by our choices of s and g, we have that

$$\Pr(X < (1+\delta)\mu) < n^{-4/3}.$$

Thus, with high probability, we avoid this bad case, as well.

We leave as an exercise to prove that our choice of s implies that $|R|$ is $O(n^{2/3})$ with high probability; hence, our analysis so far implies that, with high probability, the number of comparisons performed is $n + k + O(n^{2/3} \log n) + O(|M| \log |M|)$. Let us therefore next bound, with high probability, the size of M, which, in the general case, consists of the elements in S between u and v. To bound this, let us consider the probability that any subset, T, of $4gn/s$ contiguously-ranked elements in S would have fewer than $2g$ elements chosen to be in R, since there are $2g$ elements in R between u and v. If we can show that this occurs with low probability, then $|M| \leq 4gn/s$ with high probability. Let Y denote the number of elements in T that are chosen to be in R. Then $\mu = E[Y] = 4g$; hence, setting $\delta = 1/2$ and applying Theorem 11,

$$\Pr(Y < 2g) = \Pr(X < (1-\delta)\mu) < 2^{-\mu/8} = 2^{-g/2} = 2^{-(n^{1/3} \log^{1/2} n)/2}.$$

Thus, with high probability, by our choices of s and g, $|M| \leq 4gn/s = 4n^{2/3} \log^{1/2} n$, which implies that, with high probability, the number of comparisons performed by the algorithm is $n + k + O(n^{2/3} \log^{3/2} n)$, which establishes the theorem. □

7. Conclusion

In this paper, we have provided parameterized Chernoff bounds that have probabilities that are powers of two, and shown how to apply them for simplified analysis of some randomized algorithms. Thus, we believe that we have provided evidence that parameterizing Chernoff bounds is valuable not only for its own sake, but that it also can lead to new discoveries. As we mention above, we believe our results are just the tip of the iceberg in terms of new insights that are possible by using parameterized Chernoff bounds.

As δ grows, we move from a regime where central-limit-theorem-like estimates apply towards a regime of large deviation theory [9]. For example, an interesting question for possible future work could be to compare Cramer's large deviation theorem to Chernoff bounds as δ grows. For example, even in their transcendental exponent form, Chernoff bounds are based on the inequality $1 + \delta < \exp(\delta)$, which is tight for small δ, but less good as δ grows larger.

CRediT authorship contribution statement

Michael Dillencourt: Writing – review & editing, Writing – original draft, Formal analysis, Conceptualization. **Michael T. Goodrich:** Writing – review & editing, Writing – original draft, Funding acquisition, Formal analysis, Conceptualization. **Michael Mitzenmacher:** Writing – review & editing, Writing – original draft, Formal analysis.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Michael T. Goodrich reports financial support was provided by National Science Foundation. Michael Goodrich reports a relationship with National Science Foundation that includes: funding grants. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

This research was supported in part by the National Science Foundation under grant 2212129. Michael Mitzenmacher's work was supported in part by the National Science Foundation under grants CCF-2101140, CNS-2107078, and DMS-2023528.

References

[1] N. Alon, J.H. Spencer, The Probabilistic Method, 4th edition, John Wiley & Sons, 2016.
[2] D. Angluin, L.G. Valiant, Fast probabilistic algorithms for Hamiltonian circuits and matchings, in: 9th ACM Symposium on Theory of Computing (STOC), 1977, pp. 30–41.
[3] D.A. Barry, P.J. Culligan-Hensley, S.J. Barry, Real values of the W-function, ACM Trans. Math. Softw. 21 (2) (1995) 161–171.

[4] H. Chernoff, A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, Ann. Math. Stat. (1952) 493–507.

[5] R.M. Corless, G.H. Gonnet, D.E. Hare, D.J. Jeffrey, D.E. Knuth, On the Lambert W function, Adv. Comput. Math. 5 (1) (1996) 329–359.

[6] R.M. Corless, D.J. Jeffrey, D.E. Knuth, A sequence of series for the Lambert W function, in: Int. Symp. on Symbolic and Algebraic Computation, 1997, pp. 197–204.

[7] L. Devroye, A note on the height of binary search trees, J. ACM 33 (3) (1986) 489–498.

[8] M. Dillencourt, M.T. Goodrich, Simplified Chernoff bounds with powers-of-two probabilities, Inf. Process. Lett. (2023) 106397.

[9] R.S. Ellis, An overview of the theory of large deviations and applications to statistical mechanics, Scand. Actuar. J. 1995 (1) (1995) 97–142.

[10] R. Elsässer, T. Sauerwald, On the runtime and robustness of randomized broadcasting, Theor. Comput. Sci. 410 (36) (2009) 3414–3427.

[11] D. Eppstein, Selection and Order Statistics, Lecture Notes, 1996, https://www.ics.uci.edu/~eppstein/161/960125.html.

[12] R.W. Floyd, R.L. Rivest, Expected time bounds for selection, Commun. ACM 18 (3) (mar 1975) 165–172.

[13] A.V. Gerbessiotis, C.J. Siniolakis, A probabilistic analysis of the Floyd-Rivest expected time selection algorithm, Int. J. Comput. Math. 82 (5) (2005) 509–519.

[14] T. Hagerup, C. Rüb, A guided tour of Chernoff bounds, Inf. Process. Lett. 33 (6) (1990) 305–308.

[15] W. Hoeffding, Probability inequalities for sums of bounded random variables, J. Am. Stat. Assoc. 58 (301) (1963) 13–30.

[16] R. Iacono, J.P. Boyd, New approximations to the principal real-valued branch of the Lambert W-function, Adv. Comput. Math. 43 (2017) 1403–1436.

[17] K.C. Kiwiel, On Floyd and Rivest's SELECT algorithm, Theor. Comput. Sci. 347 (1) (2005) 214–238.

[18] M. Mitzenmacher, E. Upfal, Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis, 2nd edition, Cambridge University Press, 2017.

[19] R. Motwani, P. Raghavan, Randomized Algorithms, Cambridge University Press, 1995.

[20] B. Reed, The height of a random binary search tree, J. ACM 50 (3) (2003) 306–332.

[21] J.P. Schmidt, A. Siegel, A. Srinivasan, Chernoff–Hoeffding bounds for applications with limited independence, SIAM J. Discrete Math. 8 (2) (1995) 223–250.

[22] D. Shiu, Efficient computation of tight approximations to Chernoff bounds, Comput. Stat. (2022) 1–15.

[23] G.B. Thomas, R.L. Finney, Calculus and Analytical Geometry, 9th edition, Addison Wesley, 1996.

[24] L.G. Valiant, A scheme for fast parallel communication, SIAM J. Comput. 11 (2) (1982) 350–361.

[25] L.G. Valiant, G.J. Brebner, Universal schemes for parallel communication, in: Proceedings of the Thirteenth Annual ACM Symposium on Theory of Computing, 1981, pp. 263–277.