Title
Simplified Chernoff bounds with powers-of-two probabilities

Authors
Dillencourt, Michael
Goodrich, Michael T

Publication Date
2023-08-01

DOI
10.1016/j.ipl.2023.106397

Peer reviewed

Information Processing Letters 182 (2023) 106397

Contents lists available at ScienceDirect

Information Processing Letters

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

Simplified Chernoff bounds with powers-of-two probabilities

Michael Dillencourt, Michael T. Goodrich *

Department of Computer Science, University of California, Irvine, CA 92697, USA

ARTICLE INFO

Article history:
Received 20 September 2022
Received in revised form 17 February 2023
Accepted 2 April 2023
Available online 5 April 2023

Keywords:
Algorithms
Analysis of algorithms
Algorithm analysis
Chernoff bound
Lambert W function

ABSTRACT

In this paper, we derive simplified Chernoff bounds with powers-of-two probabilities, and we show their uses in analyzing probabilistic algorithms.

1. Introduction

Chernoff bounds [4,11] have been shown to be useful for analyzing a wide variety of different probabilistic algorithms and processes, e.g., see [1,9,12,13].

Suppose $X_1, X_2, \ldots, X_n$ are 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. In their more general (multiplicative) forms for such a random variable, X, Chernoff bounds can be stated as follows, e.g., see [1,9,12–14].

Theorem 1. *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 unwieldy to use in practice, however; hence, researchers often use other forms of the Chernoff bounds, with the following being common (see, e.g., [1,2,9,12–14]):

* Corresponding author.
E-mail addresses: dillenco@ics.uci.edu (M. Dillencourt), goodrich@uci.edu (M.T. Goodrich).

https://doi.org/10.1016/j.ipl.2023.106397

Theorem 2.

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

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

As evidence for how influential these bounds have been, we note that a 1990 paper from *Information Processing Letters*, by Hagerup and Rüb [9], which includes bounds like those in Theorem 2, has been cited over 700 times! Indeed, these bounds have become so well-known that researchers often use them without citation.

As in Theorem 2, probabilities in simplified Chernoff bounds are typically expressed as powers of Euler's number, e, whereas in Computer Science applications it is often preferred to express probabilities in terms of powers of 2, for which simplified Chernoff bounds are lacking. Indeed, some researchers apply a Chernoff bound, as in Theorem 2, and then convert the resulting probability to a power of two using the crude inequality, $2 \le e$. For example, see Elsässer and Sauerwald [8]. Of course, one can use a slightly better inequality to derive the following.

Corollary 3.

$$\Pr(X > (1+\delta)\mu) < 2^{-1.442\delta^2\mu/(2+\delta)} < 2^{-7\delta^2\mu/(10+5\delta)}, \quad \textit{for } \delta > 0, \tag{3}$$

$$\Pr(X < (1-\delta)\mu) < 2^{-1.442\delta^2\mu/2} < 2^{-7\delta^2\mu/10}, \quad \textit{for } 0 < \delta < 1. \tag{4}$$

Proof. Note that $2^{7/5} < e$, since $\log_2 e \approx 1.442695$; hence, the bounds follow immediately from Theorem 2. ■

In this paper, we are interested in simplified Chernoff bounds with powers-of-two probabilities for reasonable values of δ, as such bounds are often used in Computer Science applications. In terms of prior work, there is a notable upper-tail power-of-two Chernoff bound from a book by Mitzenmacher and Upfal [12] and the *IPL* paper by Hagerup and Rüb [9]:

Theorem 4 *([12] (p. 69) and [9]).*

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

In addition, Motwani and Raghavan [13] leave as an exercise to prove $\Pr(X > R) < 2^{-R}$, for $R \ge 2e\mu$, which is a slightly better condition, since $2e \approx 5.43656$. Although this and the power-of-two Chernoff bound of Theorem 4 are useful, we show below that $\Pr(X > R) < 2^{-R}$, for $R \ge 4.5\mu$, which can lead to better analyses for randomized algorithms. Indeed, in this paper, we derive a number of such simplified Chernoff bounds with powers-of-two probabilities, for both upper and lower tails, for reasonable values of δ. We also mention some applications of our simplified powers-of-two Chernoff bounds, but these are just the tip of the iceberg in terms of improved analyses of algorithms that are possible, e.g., given that the *IPL* paper by Hagerup and Rüb [9] has been cited over 700 times.

2. The Lambert W function

Some of our proofs make use of the Lambert W function; hence, before we derive our simplified powers-of-two Chernoff bounds, let us first review this function. The Lambert W function is defined by the rule that $W(z) = w$ iff w satisfies the equation

$$we^w = z,$$

e.g., see Corless, Gonnet, Hare, Jeffrey, and Knuth [5] or Corless, Jeffrey, and Knuth [6]. Technically, W is not a function. Hence its real-valued expression is partitioned into two branches: $W_0(x)$, 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 $(-\frac{1}{e}, -1)$. $W_0(ye^y) = y$ for $y \ge -1$, and $W_{-1}(ye^y) = y$ for $y \le -1$.

The Lambert W function has many applications, including characterizing the number of unrooted trees [5]. It cannot be expressed in terms of elementary functions; hence, evaluating it typically requires one to use a numerical algorithm, e.g., see [3].

3. Improved powers-of-two Chernoff bounds

In this section, we derive simplified Chernoff bounds with powers-of-two probabilities.

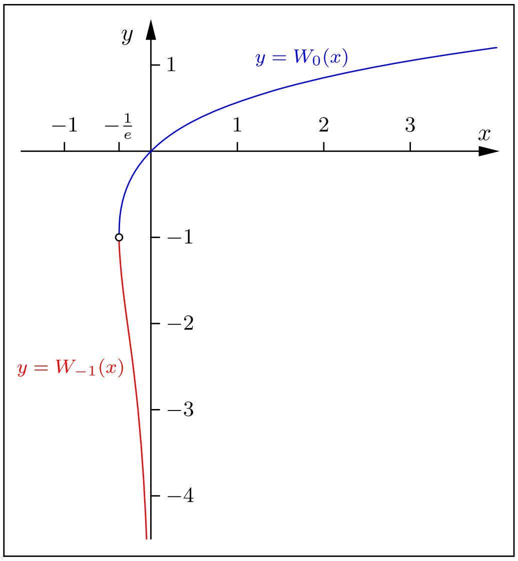

Fig. 1. The two real branches of the Lambert W function. Image Copyright © 2022 Michael Dillencourt; used with permission.

3.1. Upper-tail bounds

We begin with some upper-tail bounds. The first is a strict improvement of Theorem 4.

Theorem 5.

$$\Pr(X > R) < 2^{-R}, \quad \textit{for } R \geq 4.5\mu.$$

Proof. We actually show that $\Pr(X > R) < 2^{-R}$, for $R \geq 4.31107\mu > -\mu/W_0(-1/2e)$, where $W_0(x)$ is the principal branch of the Lambert W function. From the general form of the Chernoff bound of Theorem 1, taking $R = (1+\delta)\mu$,

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

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

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

That is,

$$xe^{-\frac{x-1}{x}} = 2.$$

Putting this into the form of the Lambert W function definition, let $u = -(x-1)/x$, so $x = 1/(1+u)$. The equation becomes

$$\left(\frac{1}{1+u}\right)e^u = 2,$$

which can be rewritten as

$$-(1+u)e^{-(1+u)} = -\frac{1}{2e}.$$

This equation has $u = -W_0(-1/2e) - 1$ as a solution. After back-substituting the solution is $x = -1/W_0(-1/2e)$. Numerically, $-1/W_0(-1/2e) \approx -1/(-0.23196) \approx 4.31107$, which establishes the bound for R. ■

We can also establish the following general bounds.

Theorem 6. *The bound*

$$\Pr(X > (1+\delta)\mu) < 2^{-\alpha\mu}, \tag{5}$$

holds:

1. *For fixed* $\delta > 0$ *when*

$$\alpha \le \log_2 \left(\frac{(1+\delta)^{1+\delta}}{e^{\delta}} \right). \tag{6}$$

2. *For fixed* $\alpha > 0$ *when*

$$\delta \ge e^{W_0\left(\frac{\alpha \ln 2 - 1}{e}\right)+1} - 1. \tag{7}$$

Proof. By Theorem 1, (5) holds whenever we have:

$$2^{-\alpha\mu} \ge \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}} \right)^{\mu}. \tag{8}$$

Part 1 follows from the observation that if (6) holds, so does (8). To establish part 2, we fix δ and determine the conditions for which (8) holds. (8) holds when

$$\left(\frac{1+\delta}{e} \right)^{1+\delta} \ge \frac{2^{\alpha}}{e}.$$

Since $\ln x$ and x/e are both monotone increasing functions of x for positive x, this is equivalent to

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

Since the left-hand side is of the form xe^x where $x = \ln((1+\delta)/e)$, and since $\ln((1+\delta)/e) > -1$ for any positive δ, we can rewrite the last equation as

$$\ln \left(\frac{1+\delta}{e} \right) \ge W_0 \left(\frac{\alpha \ln 2 - 1}{e} \right),$$

from which (7) follows. ■

We also have the following theorem, which provides a bound for modest values of δ.

Theorem 7.

$$\Pr(X > (1+\delta)\mu) < 2^{-\delta\mu}, \quad \text{for } \delta > 2.20603. \tag{9}$$

Proof. From the general form of the Chernoff bound of Theorem 1,

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

For this probability to be at most $2^{-\delta\mu}$, we need $2^{-\delta}(1+\delta)^{1+\delta} \ge e^{\delta}$. This can be rewritten as

$$\left(\frac{1+\delta}{2e} \right)^{1+\delta} \ge \frac{1}{2e}. \tag{10}$$

Taking both sides to the $1/(2e)$ power, and taking the log of both sides yields

$$\left(\frac{1+\delta}{2e} \right) \ln \left(\frac{1+\delta}{2e} \right) \ge \left(\frac{1}{2e} \right) \ln \left(\frac{1}{2e} \right). \tag{11}$$

We can see by inspection that one breakpoint for (11) is $\delta = 0$, which is the trivial value for which equality holds in (10). Since $\ln(1/(2e)) < -1$, this corresponds to choosing the non-principal branch of the Lambert function. Choosing the principal branch, we see that equality occurs in (11) when

$$W_0 \left(\frac{-\ln 2 - 1)}{2e} \right) = \ln \left(\frac{1+\delta}{2e} \right).$$

This produces the solution

$$\delta = 2e \cdot e^{W_0\left(\frac{-\ln 2 - 1}{2e}\right)} - 1 \approx 2e \cdot e^{W_0(0.31144)} - 1 \approx 2e \cdot e^{-0.52811} - 1 \approx 2.20603. \quad ■$$

We can use Theorem 6 to derive the following specific upper-tail powers-of-two Chernoff bounds for smaller values of δ, all of which are tighter than the bounds of Corollary 3.

Corollary 8.

$$\Pr(X > (1+\delta)\mu) < 2^{-\mu}, \quad \textit{for } \delta \geq 1.4, \tag{12}$$

$$\Pr(X > (1+\delta)\mu) < 2^{-0.557\mu} < 2^{-5\mu/9}, \quad \textit{for } \delta \geq 1, \tag{13}$$

$$\Pr(X > (1+\delta)\mu) < 2^{-0.266\mu} < 2^{-\mu/4}, \quad \textit{for } \delta \geq 2/3, \tag{14}$$

$$\Pr(X > (1+\delta)\mu) < 2^{-0.156\mu} < 2^{-3\mu/20}, \quad \textit{for } \delta \geq 1/2, \tag{15}$$

$$\Pr(X > (1+\delta)\mu) < 2^{-0.072\mu} < 2^{-\mu/14}, \quad \textit{for } \delta \geq 1/3, \tag{16}$$

$$\Pr(X > (1+\delta)\mu) < 2^{-0.0417\mu} < 2^{-\mu/24}, \quad \textit{for } \delta \geq 1/4. \tag{17}$$

Proof. To derive (12), we set $\alpha = 1$ and use part 2 of Theorem 6. By (7), the minimum value of δ is

$$e^{W_0\left(\frac{\ln 2-1}{e}\right)+1} - 1 \approx e^{W_0(-0.11288)+1} - 1 \approx e^{-0.128337+1} - 1 \approx 1.39088 < 1.4.$$

To derive (13), we set $\delta = 1$ and use part 1 of Theorem 6. By (6), the maximum value of α is

$$\log_2\left(\frac{2^2}{e^2}\right) \approx 0.557.$$

The derivations of (14) through (17) are similar to that of (13). ■

3.2. Lower-tail bounds

We also derive lower-tail Chernoff bounds that improve the Chernoff bounds of Corollary 3. We first prove the following analog of Theorem 6.

Theorem 9. *The bound*

$$\Pr(X < (1-\delta)\mu) < 2^{-\beta\mu}, \tag{18}$$

holds:

1. *For fixed δ with $0 < \delta < 1$ when*

$$\beta \leq \log_2\left(e^{\delta}(1-\delta)^{1-\delta}\right). \tag{19}$$

2. *For fixed $\beta > 0$ when*

$$1 - e^{W_{-1}\left(\frac{\beta\ln 2-1}{e}\right)+1} \leq \delta < 1. \tag{20}$$

Proof. By Theorem 1, (18) holds whenever we have:

$$2^{-\beta\mu} \geq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu} \tag{21}$$

Part 1 follows from the observation that if (19) holds, so does (21). To establish part 2, we fix β and determine the values of δ between 0 and 1 for which (21) holds. (21) holds when

$$\left(\frac{1-\delta}{e}\right)^{1-\delta} \geq \frac{2^{\beta}}{e}. \tag{22}$$

Since the functions $\ln x$ and x/e are both monotone increasing functions of x for positive x and since both quantities in (22) are less than 1, this is equivalent to

$$\left(\frac{1-\delta}{e}\right)\ln\left(\frac{1-\delta}{e}\right) \leq \frac{\beta\ln 2 - 1}{e}.$$

Since the left-hand side is of the form xe^x where $x = \ln((1-\delta)/e)$, and since $\ln((1-\delta)/e) < -1$ for any positive δ, we can rewrite the last equation as

$$\ln\left(\frac{1-\delta}{e}\right) \le W_{-1}\left(\frac{\beta\ln 2-1}{e}\right),$$

from which (20) follows. ■

Corollary 10.

$$\Pr(X < (1-\delta)\mu) < 2^{-\mu}, \quad \text{for } 0.9099 \le \delta < 1. \tag{23}$$

Proof. To derive (23), we set $\beta = 1$ and use part 2 of Theorem 9. By (20), the minimum value of δ is

$$1 - e^{W_{-1}\left(\frac{\ln 2-1}{e}\right)+1} \approx 1 - e^{W_{-1}(-0.11288)+1} - 1 \approx 1 - e^{-3.40737+1} \approx 0.9099. \quad ■$$

If we are interested in bounds for slightly smaller values of δ, we can use the following theorem.

Theorem 11.

$$\Pr(X < (1-\delta)\mu) < 2^{-\delta\mu}, \quad \text{for } 0.8687 \le \delta < 1. \tag{24}$$

Proof. From the general form of the Chernoff bound of Theorem 1,

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

For this probability to be at most $2^{-\delta\mu}$, we need $e^{\delta}(1-\delta)^{1-\delta} \ge 2^{\delta}$. This can be rewritten as

$$\left(\frac{2(1-\delta)}{e}\right)^{1-\delta} \ge \frac{2}{e}. \tag{25}$$

Taking both sides to the $2/e$ power, and taking the log of both sides yields

$$\left(\frac{2(1-\delta)}{e}\right)\ln\left(\frac{2(1-\delta)}{e}\right) \ge \left(\frac{2}{e}\right)\ln\left(\frac{2}{e}\right). \tag{26}$$

We can see by inspection that one breakpoint for (26) is $\delta = 0$, which is the trivial value for which equality holds in (25). Since $\ln(2/e) > -1$, this corresponds to choosing the principal branch of the Lambert function. Choosing the non-principal branch, we see that equality occurs in (26) when

$$W_{-1}\left(\frac{2(\ln 2-1)}{e}\right) = \ln\left(\frac{2(1-\delta)}{e}\right).$$

This produces the solution

$$\delta = 1 - \frac{e^{W_{-1}\left(\frac{2(\ln 2-1)}{e}\right)+1}}{2} \approx 1 - \frac{e^{W_{-1}(-0.2257)+1}}{2} \approx 1 - \frac{e^{-2.3372+1}}{2} \approx 0.8687. \quad ■$$

We can also derive the following specific lower-tail powers-of-two Chernoff bounds for smaller values of δ, all of which are tighter than the bounds of Corollary 3.

Corollary 12. *Suppose $\delta < 1$. Then*

$$\Pr(X < (1-\delta)\mu) < 2^{-0.771\mu} < 2^{-3\mu/4}, \quad \text{for } \delta \ge 5/6, \tag{27}$$
$$\Pr(X < (1-\delta)\mu) < 2^{-0.582\mu} < 2^{-5\mu/9}, \quad \text{for } \delta \ge 3/4, \tag{28}$$
$$\Pr(X < (1-\delta)\mu) < 2^{-0.433\mu} < 2^{-3\mu/7}, \quad \text{for } \delta \ge 2/3, \tag{29}$$
$$\Pr(X < (1-\delta)\mu) < 2^{-0.221\mu} < 2^{-\mu/5}, \quad \text{for } \delta \ge 1/2, \tag{30}$$
$$\Pr(X < (1-\delta)\mu) < 2^{-0.09092\mu} < 2^{-\mu/11}, \quad \text{for } \delta \ge 1/3, \tag{31}$$
$$\Pr(X < (1-\delta)\mu) < 2^{-0.049\mu} < 2^{-\mu/21}, \quad \text{for } \delta \ge 1/4. \tag{32}$$

Proof. To derive (27), we set $\delta = 5/6$ and use part 1 of Theorem 9. By (19), the maximum value of β is

$$\log_2\left(e^{5/6}(1/6)^{1/6}\right) \approx 0.7714.$$

The derivations of (28) through (32) are similar to that of (27). ■

4. Applications

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

Hassin and Peleg [10] study a probabilistic local polling process, examine its properties, and propose its use in the context of distributed network protocols for achieving consensus. Their analysis uses Theorem 4 to show that a parallel random-walk process will succeed with half of the pairs of random walks meeting in $41M$ expected steps, where M is the maximum expected meeting time for two walks. Substituting Theorem 5 in their analysis improves the expected number of steps for half of the pairs meeting to $24M$.

Diks and Pelc [7] present an algorithm to exchange values between all fault-free nodes in an n-node network where nodes and links fail with constant probabilities, basing their analysis, in part, on Theorem 4. Substituting Theorem 5 in their analysis improves the constant factor in their analysis and/or the probability of failure that their algorithm tolerates.

Elsässer and Sauerwald [8] study a randomized broadcasing protocol. Their analysis uses Theorem 2, with $\delta = 5/6$, and the crude inequality $2 < e$ to bound the failure probability of their algorithm to be at most $1/n$. Simply substituting Theorem 12 in their analysis improves their failure probability to $n^{-3.5}$.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Michael Goodrich reports financial support was provided by National Science Foundation (2212129). Michael Goodrich reports a relationship with National Science Foundation (2212129) that includes: funding grants.

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.

References

[1] N. Alon, J.H. Spencer, The Probabilistic Method, 4/e 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, The Annals of Mathematical Statistics (1952) 493–507.
[5] R.M. Corless, G.H. Gonnet, D.E. Hare, D.J. Jeffrey, D.E. Knuth, On the Lambert W function, Advances in Computational Mathematics 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 Comp., 1997, pp. 197–204.
[7] K. Diks, A. Pelc, Efficient gossiping by packets in networks with random faults, SIAM Journal on Discrete Mathematics 9 (1) (1996) 7–18.
[8] R. Elsässer, T. Sauerwald, On the runtime and robustness of randomized broadcasting, Theoretical Computer Science 410 (36) (2009) 3414–3427.
[9] T. Hagerup, C. Rüb, A guided tour of Chernoff bounds, Information Processing Letters 33 (6) (1990) 305–308.
[10] Y. Hassin, D. Peleg, Distributed probabilistic polling and applications to proportionate agreement, Information and Computation 171 (2) (2001) 248–268.
[11] W. Hoeffding, Probability inequalities for sums of bounded random variables, Journal of the American Statistical Association 58 (301) (1963) 13–30.
[12] M. Mitzenmacher, E. Upfal, Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis, 2/e edition, Cambridge University Press, 2017.
[13] R. Motwani, P. Raghavan, Randomized Algorithms, Cambridge University Press, 1995.
[14] D. Shiu, Efficient computation of tight approximations to Chernoff bounds, Computational Statistics (2022) 1–15.