Efficient computation of tight approximations to Chernoff bounds

Daniel Shiu

Received: 5 October 2021 / Accepted: 14 March 2022 / Published online: 4 April 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Chernoff bounds are a powerful application of the Markov inequality to produce strong bounds on the tails of probability distributions. They are often used to bound the tail probabilities of sums of Poisson trials, or in regression to produce conservative confidence intervals for the parameters of such trials. The bounds provide expressions for the tail probabilities that can be inverted for a given probability/confidence to provide tail intervals. The inversions involve the solution of transcendental equations and it is often convenient to substitute approximations that can be exactly solved e.g. by the quadratic equation. In this paper we introduce approximations for the Chernoff bounds whose inversion can be exactly solved with a quadratic equation, but which are closer approximations than those adopted previously.

Keywords Chernoff bounds · Tail distributions · Concentration inequalities · Poisson trials

1 Introduction

Chernoff bounds (Dubhashi and Panconesi 2009; Mitzenmacher and Upfal 2017) produce exponentially decreasing concentration inequalities that are both rigorous and commensurable with the central limit theorem. They are very effective at bounding the tail probabilities of binomial distributions, or more generally sums of Poisson trials. The bounds produced are much stronger than those provided by generic bounds such as Chebyshev’s inequality. In their multiplicative form, for a given $\delta$ Chernoff bounds provide tight bounds on the probability that random variables deviate from their expectation by factors more than $(1 + \delta)$ or less than $(1 - \delta)$. Such bounds are used in the analysis of randomised algorithms (Dubhashi and Panconesi 2009), efficient network routing (Mitzenmacher and Upfal 2017), set balancing and parameter...
estimation in quantum communications (Zhang et al. 2017). For many applications the
reverse calculation is required: given a bound on probability to find a factor so that the
probability of deviation beyond that factor is less than the bound. This motivates the
derivation of looser bounds whose inversion is a less complex calculation. Accordingly,
simpler bounds whose inversions require only a simple arithmetic combination of
logarithms and square roots are often used in practice.

In this paper we demonstrate bounds for sums of Poisson trials that again can be
inverted with closed form expressions involving logarithms and square roots. However,
these bounds are closer to the original Chernoff expression. We develop these bounds
by taking quadratic Padé approximations to the logarithm of the bound. We use slightly
different derivations for prediction and regression applications so that a more accurate
multiplicative form of the regressed confidence interval is produced.

The motivation for a more accurate, computationally tractable approximation arises
principally from the field of security. The pre-existing approximations work well for
confidence intervals where the tail probabilities are regularly used values such as 0.1,
0.05 or 0.01. The accuracy of the approximations decreases as the tail probability
decreases. In the field of cryptographic security however, if rare events can compro-
mise data then the probability of such events must be kept to extremely low values.
For example, in section 5.3 of their analysis of decryption failures in NIST candidate
post-quantum algorithms (Bindel and Schanck 2020), Bindel and Schanck describe
a tail probability greater than $2^{-60}$ as “concerning”. Similarly, in the area of quan-
tum communications, when Pirandola describes worst case parameter estimation in
section III.D of his analysis of quantum communication limits (Pirandola 2021), tail
probability values of $10^{-10}$, $10^{-17}$ and even $10^{-43}$ are quoted. Tail probabilities such
as these are used with Chernoff bounds in parameter estimation methods such as Zhang
et al. (2017).

A simple example of the use of the new bounds is in the confident measurement
of entropy from noise sources. Suppose that we have a physical noise source that
produces random i.i.d. bits according to a Bernoulli random variable which produces
a 1 with (unknown) probability $p$. If we require a confidence of $2^{-40}$ in the minimum
entropy of the output, we will need a similarly confident estimate of $p$. If we take
1000 output samples and detect 686 outputs of 1, the old approximations allow us
to confidently predict that $p < 0.959$ whereas the new approximations allow us
to confidently predict that $p < 0.9004$. This leads to an improved, highly confident
estimate of the minimum entropy by a factor $\log(0.9004) / \log(0.959)$ which represents
an improvement of almost 250% in the rate of generation of randomness.

2 Approximating tail distributions

We begin with the predictive application for sums of independent Poisson trials $X_i$
where the sum $\sum X_i$ has expectation $\mu$. We have (see for example Mitzenmacher and
Upfal 2017, Theorems 4.4 and 4.5) the Chernoff bound tail probabilities:
\[ \mathbb{P}\left( \sum_{i} X_i \geq (1 + \delta) \mu \right) \leq \exp\left((\delta - (1 + \delta) \log(1 + \delta)) \mu\right) \]

\[ \mathbb{P}\left( \sum_{i} X_i \leq (1 - \delta) \mu \right) \leq \exp\left((-\delta - (1 - \delta) \log(1 - \delta)) \mu\right). \]

If we are given an upper bound tail probability, \( \gamma \) and we might be required to find \( \delta_U \) such that \( \mathbb{P}(\sum X_i \geq (1 + \delta U) \mu) < \gamma \) or such that \( \mathbb{P}(\sum X_i \leq (1 - \delta_L) \mu) < \gamma \). We could solve

\[ \delta_U - (1 + \delta_U) \log(1 + \delta) = \frac{\log \gamma}{\mu} \] (1)

or

\[ -\delta_L - (1 - \delta_L) \log(1 - \delta) = \frac{\log \gamma}{\mu} \] (2)

numerically (e.g. by iterative methods such as binary search or Newton’s method).

In practice approximations \((1 - \delta) \log(1 - \delta) \geq -\delta + \delta^2/2\) and \(\log(1 + \delta) \geq 2\delta/(2 + \delta)\) valid for \(0 \leq \delta < 1\) are used to provide the more wieldy bounds

\[ \mathbb{P}\left( \sum_{i} X_i \geq (1 + \delta) \mu \right) \leq \exp\left(-\frac{\delta^2 \mu}{2 + \delta}\right) \] (3)

\[ \mathbb{P}\left( \sum_{i} X_i \leq (1 - \delta) \mu \right) \leq \exp\left(-\frac{\delta^2 \mu}{2}\right). \] (4)

For our tail probability \( \gamma \) we can use these looser expressions to form the quadratic equations

\[ \delta_U^2 + \left(\frac{\log \gamma}{\mu}\right) \delta_U + 2 \left(\frac{\log \gamma}{\mu}\right) = 0 \]

\[ \delta_L^2 + 2 \left(\frac{\log \gamma}{\mu}\right) = 0 \]

and evaluate the closed form expressions

\[ \delta_U = \frac{-\log \gamma + \sqrt{(\log \gamma/\mu)^2 - 8 \log \gamma}}{2} \] (5)

\[ \delta_L = \sqrt{-\frac{2 \log \gamma}{\mu}} \] (6)

and conclude that \( \mathbb{P}(\sum X_i \geq (1 + \delta U) \mu) < \gamma \) and \( \mathbb{P}(\sum X_i \leq (1 - \delta L) \mu) < \gamma \). If a two-tailed bound is required, we can note that \( \delta_U \geq \delta_L \) so that \( \mathbb{P}((1 - \delta U) \mu \leq \sum X_i \leq (1 + \delta U) \mu) < 1 - 2\gamma \).
The inequalities \((1 - \delta) \log(1 - \delta) \geq -\delta + \delta^2/2\) and \(\log(1 + \delta) \geq 2\delta/(2 + \delta)\) are quadratic Padé approximations to \(\log(1 \pm \delta)\), but we can do better with approximations to the functions

\[
\delta - (1 + \delta) \log(1 + \delta) = -\frac{\delta^2}{2} + \frac{\delta^3}{6} - \frac{\delta^4}{12} + \frac{\delta^5}{20} - \frac{\delta^6}{30} - \cdots
\]

and

\[
- \delta - (1 - \delta) \log(1 - \delta) = -\frac{\delta^2}{2} - \frac{\delta^3}{6} - \frac{\delta^4}{12} - \frac{\delta^5}{20} - \frac{\delta^6}{30} - \cdots
\]

We see that

\[
\frac{-3\delta^2}{6 + 2\delta} = -\frac{\delta^2}{2} + \frac{\delta^3}{6} - \frac{\delta^4}{12} + \frac{\delta^5}{20} - \frac{\delta^6}{30} - \cdots
\]

and

\[
\frac{-9\delta^2}{18 - 6\delta - \delta^2} = -\frac{\delta^2}{2} - \frac{\delta^3}{6} - \frac{\delta^4}{12} - \frac{\delta^5}{27} - \frac{11\delta^6}{648} - \cdots
\]

It is easy to confirm that for \(\delta \in (0, 1)\) we have

\[
\delta - (1 + \delta) \log(1 + \delta) < -\frac{3\delta^2}{6 + 2\delta}
\]

and

\[
-\delta - (1 - \delta) \log(1 - \delta) < -\frac{9\delta^2}{18 - 6\delta - \delta^2},
\]

so that

\[
P\left(\sum_i X_i \geq (1 + \delta)\mu\right) \leq \exp\left(\frac{-3\delta^2}{6 + 2\delta}\right)
\]

and

\[
P\left(\sum_i X_i \leq (1 - \delta)\mu\right) \leq \exp\left(\frac{-9\delta^2}{18 - 6\delta - \delta^2}\right).
\]

Note that the difference between the power series in (7) and (9) is \(O(\delta^4)\) and the difference between the power series in (8) and (10) is \(O(\delta^5)\). For comparison, the bounds in (3) and (4) introduce a \(O(\delta^3)\) difference into the exponent. We also see that if we restrict to identically distributed trials (i.e. Bernoulli rather than Poisson), Eq. (11) can be rearranged to an estimate previously derived using Bernstein inequalities (Bernstein 1924). For our tail probability \(\gamma\) we can use these more accurate expressions to form the quadratic equations

\[
3\delta_U^2 + 2\left(\frac{\log \gamma}{\mu}\right) \delta_U + 6\left(\frac{\log \gamma}{\mu}\right) = 0
\]

\[
\left(9 - \frac{\log \gamma}{\mu}\right) \delta_L^2 - 6\left(\frac{\log \gamma}{\mu}\right) \delta_L + 18\left(\frac{\log \gamma}{\mu}\right) = 0
\]
and evaluate the following closed form expressions which are sharper alternatives to (5) and (6)

\[
\delta_U = -\left(\frac{\log \gamma}{\mu}\right) + \sqrt{\left(\frac{\log \gamma}{\mu}\right)^2 - 18 \left(\frac{\log \gamma}{\mu}\right)}
\]

\[
\delta_L = \frac{3 \left(\frac{\log \gamma}{\mu}\right) + \sqrt{9 \left(\frac{\log \gamma}{\mu}\right)^2 - 18 \left(\frac{\log \gamma}{\mu}\right) (9 - \frac{\log \gamma}{\mu})}}{9 - \frac{\log \gamma}{\mu}}.
\]

We have now proven the following:

**Theorem 1** Let \( X_1, \ldots, X_n \) be independent Poisson trials. Let \( X = \sum_{i=1}^n X_i \) and \( \mu = \mathbb{E}[X] \). Let \( 0 < \gamma < 1 \) be a fixed tail probability and write \( \beta = (\log \gamma)/\mu \). Let

\[
\delta_U = -\beta + \sqrt{\beta^2 - 18 \beta}
\]

\[
\delta_L = 3 \left(\frac{\beta + \sqrt{\beta^2 - 2 \beta (9 - \beta)}}{9 - \beta}\right).
\]

Then

\[
\mathbb{P}(X \geq (1 + \delta_U)\mu) < \gamma
\]

and

\[
\mathbb{P}(X \leq (1 - \delta_L)\mu) < \gamma.
\]

### 3 Approximating confidence intervals

For our regression application, we have an observed value \( \hat{\mu} \) of \( \sum X_i \) and a target confidence \( \gamma \). Our goal is to identify a range of possible underlying \( \mu \) values such that the likelihood of our observation \( \hat{\mu} \) for \( \mu \) in that range is less than or equal to \( \gamma \). The complement of the range then provides a conservative confidence interval for \( \mu \) with confidence at least \( 1 - \gamma \).

We first develop Chernoff bounds of a slightly different form. Starting from the generic Markov bound applied to \( e^{tX} \) we have

\[
\mathbb{P}(X \leq a) \leq \min_{t < 0} \frac{\mathbb{E}[e^{tX}]}{e^{ta}}, \quad \mathbb{P}(X \leq a) \geq \min_{t > 0} \frac{\mathbb{E}[e^{tX}]}{e^{ta}}.
\]

Following the usual argument, we note that \( \mathbb{E}[e^{tX_i}] = 1 + p_i (e^t - 1) \leq \exp(p_i (e^t - 1)) \) and so \( \mathbb{E}[e^{tX}] \leq \exp((e^t - 1)\mu) \). Then with \( t = -\log(1 + \delta) \) and \( t = -\log(1 - \delta) \) we have
\[ P \left( \sum_i X_i \leq \frac{\mu}{1 + \delta} \right) \leq \exp \left( \left( -\frac{\delta}{(1 + \delta)} + \frac{\log(1 + \delta)}{(1 + \delta)} \right) \mu \right) \]

and

\[ P \left( \sum_i X_i \geq \frac{\mu}{1 - \delta} \right) \leq \exp \left( \left( \frac{\delta}{(1 - \delta)} + \frac{\log(1 - \delta)}{(1 - \delta)} \right) \mu \right). \]

Using the above bounds, if we let \( \mu_\ell \) be the value for which \( \hat{\mu} = \mu_\ell / (1 + \delta_\ell) \) where \( \delta_\ell \) is a solution to

\[ \exp \left( (\delta_\ell + \log(1 + \delta_\ell)) \hat{\mu} \right) = \gamma \]

then by monotonicity, the interval \(( -\infty, \mu_\ell ] \) is a suitable range of exceptional \( \mu \) values and \(( \mu_\ell, \infty) \) is a suitable conservative confidence interval. For a conservative confidence interval that gives an upper bound for \( \mu \), we can, by a similar process, find the \( \mu_u = \hat{\mu}(1 - \delta_u) \) where

\[ \exp \left( (\delta_u + \log(1 - \delta_u)) \hat{\mu} \right) = \gamma \]

and develop the conservative confidence interval \(( -\infty, \mu_u ] \). We can even combine our calculation for the conservative confidence interval \(( \mu_\ell, \mu_u ) \) in which we would have confidence at least \( 1 - 2\gamma \).

As in the previous section, the Eqs. (14) and (13) can be solved numerically by iterative methods such as binary search or Newton’s method. However, we seek an expression that can be solved using the quadratic formula.

For a lower bound for \( \mu \) with level of confidence \( 1 - \gamma \) with \( 0 < \gamma < 1 \), we therefore aim to identify the values \( \delta \) such that

\[ (\delta + \log(1 - \delta)) \hat{\mu} \leq \log \gamma. \]

By monotonicity, for \( 0 < \delta < 1 \), identifying the value where equality is attained proves the bound for all greater values. For a similarly confident upper bound we need to identify the values \( \delta \) such that

\[ (-\delta + \log(1 + \delta)) \hat{\mu} \leq \log \gamma \]

and again, identifying the value where equality is attained proves the bound for all greater values. Although this regression formula is not typically quoted in expositions on Chernoff bounds, one can use the standard approximations \( \log(1 - \delta) < -\delta - \delta^2 / 2 \) and \( \log(1 + \delta) < (\delta^2 + 2\delta)/(2 + 2\delta) \) for \( 0 < \delta < 1 \) to develop bounds analogous to (3) and (4):

\[ -\delta + \log(1 + \delta) \leq \frac{-\delta^2}{2 + 2\delta}. \]

\[ \delta + \log(1 - \delta) \leq \frac{-\delta^2 \mu}{2}. \]
Again though we can do better. For $0 < \delta < 1$ we have the Padé approximation to $-\delta + \log(1 + \delta)$

$$-\delta + \log(1 + \delta) < \frac{-3\delta^2}{6 + 4\delta}.$$ 

and so if we take $\delta_U$ to be the positive root of

$$3\delta_U^2 + 4\log \frac{\gamma}{\hat{\mu}} \delta_U + 6\log \frac{\gamma}{\hat{\mu}}$$

then we have

$$(-\delta_U + \log(1 + \delta_U))\hat{\mu} \leq \log \gamma.$$ 

It follows that $\mu \geq (1 + \delta_U)\hat{\mu}$ with probability at most $\gamma$ so that $\mu < (1 + \delta_U)\hat{\mu}$ with probability at least $1 - \gamma$. We note the power series expansions for $0 < \delta < 1$

$$-\delta + \log(1 + \delta) = -\frac{\delta^2}{2} + \frac{\delta^3}{3} - \frac{\delta^4}{4} + \frac{\delta^5}{5} - \frac{\delta^6}{6} - \cdots$$

$$\frac{-3\delta^2}{6 + 4\delta} = -\frac{\delta^2}{2} + \frac{\delta^3}{3} - \frac{2\delta^4}{9} + \frac{4\delta^5}{27} - \frac{8\delta^6}{81} - \cdots$$

so that our approximation differs by $O(\delta^4)$.

Likewise we also have the quadratic Padé approximation to $\delta + \log(1 - \delta)$

$$\delta + \log(1 - \delta) < \frac{-9\delta^2}{18 - 12\delta - \delta^2} \tag{19}$$

From (19) we conclude that if $\delta_L$ is taken to be the positive root of the equation

$$\left(9 - \frac{\log \gamma}{\hat{\mu}}\right)\delta_L^2 - 12\frac{\log \gamma}{\hat{\mu}}\delta_L + 18\log \frac{\gamma}{\hat{\mu}} = 0$$

then we have

$$(\delta_L + \log(1 - \delta_L))\hat{\mu} < \log \gamma.$$ 

It follows that $\mu \leq (1 - \delta_L)\hat{\mu}$ with probability at most $\gamma$ so that $\mu > (1 - \delta_L)\hat{\mu}$ with probability at least $1 - \gamma$. Again by considering the power series expansions for $0 < \delta < 1$

$$\delta + \log(1 - \delta) = -\frac{\delta^2}{2} - \frac{\delta^3}{3} - \frac{\delta^4}{4} - \frac{\delta^5}{5} - \frac{\delta^6}{6} - \cdots$$

$$\frac{-9\delta^2}{18 - 12\delta - \delta^2} = -\frac{\delta^2}{2} - \frac{\delta^3}{3} - \frac{5\delta^5}{27} - \frac{89\delta^6}{648} - \cdots$$
so that our approximation differs by $O(\delta^5)$.

The two estimates can be combined into a two-ended confidence interval allowing us to conclude that $\mu \in ((1 - \delta_L)\hat{\mu}, (1 + \delta_U)\hat{\mu})$ with probability at least $1 - 2\gamma$. Alternatively, if a symmetric expression is desired, we note that for $0 < \delta < 1$ we have

$$-\delta + \log(1 + \delta) < -\frac{3\delta^2}{6 + 4\delta} < \delta + \log(1 - \delta) < -\frac{9\delta^2}{18 - 12\delta - \delta^2}$$

and so by monotonicity $\delta_U > \delta_L$ and the interval $\mu \in ((1 - \delta_U)\hat{\mu}, (1 + \delta_U)\hat{\mu})$ can be used with confidence at least $1 - 2\gamma$.

**Theorem 2** Let $X_1, \ldots, X_n$ be independent Poisson trials. Let $X = \sum_{i=1}^n X_i$ and suppose that we have a sample $\hat{\mu}$ from $X$. Let $0 < \gamma < 1$ be a fixed bound on confidence and write $\beta = (\log \gamma)/\hat{\mu}$. Let

$$\delta_U = -2\beta + \sqrt{4\beta^2 - 18\beta}$$
$$\delta_L = 6\beta + \sqrt{36\beta^2 - 18\beta(9 - \beta)}.$$  

Then with confidence at least $1 - \gamma$ we can say

$$\mathbb{E}[X] < (1 + \delta_U)\hat{\mu}$$

and similarly with confidence at least $1 - \gamma$ we can say

$$\mathbb{E}[X] > (1 - \delta_L)\hat{\mu}.$$  

**4 Numerical examples**

We consider examples using Bernoulli trials which are a frequent use of such bounds. Suppose that we have $\mathbb{P}(X_i = 1) = 0.0002$ and that we run $1,000,000$ trials. We have $\mu = 200$. We consider tail probabilities $\gamma$ of $0.1$, $0.05$, $0.01$, $2.0e-9$ (corresponding to “six sigma”), and $5.421e-20$ (corresponding to a probability of $2^{-64}$ which is relevant to failure rates in cryptography). For each $\gamma$ we compute $\delta_U$ and $\delta_L$ using the exact transcendental Chernoff formulae, the old quadratic formulae and the new quadratic formulae of this paper. Solutions are given to four significant figures in Table 1. For context we also give the relative error between our approximations and the bound derived from the transcendental formula. Specifically, for $\delta_U$ an exact solution to one of the transcendental equations (1), (16) and $\tilde{\delta}_U$ an approximate solution to these

---

1 A feature of Chernoff bounds is that they only depend on the (expected) number of observations rather than making any explicit use of the number of trials. This means that the same $\delta$ values will be returned for any example where $\mu = 200$ e.g. 1000 trials with $\mathbb{P}(X_i = 1) = 0.2$.  

© Springer
Table 1 Comparison of tail bounds

| γ         | Exact δ_U | Exact δ_L | Old δ_U  | Old δ_L  | New δ_U  | New δ_L  |
|-----------|------------|-----------|----------|----------|----------|----------|
| 0.1       | 0.1555     | 0.1479    | 0.1576   | 0.1517   | 0.1556   | 0.1479   |
| 0.05      | 0.1780     | 0.1680    | 0.1807   | 0.1731   | 0.1781   | 0.1680   |
| 0.01      | 0.2221     | 0.2068    | 0.2264   | 0.2146   | 0.2224   | 0.2068   |
| 0.000000002 | 0.4798   | 0.4127    | 0.5004   | 0.4476   | 0.4822   | 0.4133   |
| 5.421e–20 | 0.7365     | 0.5870    | 0.7861   | 0.6660   | 0.7441   | 0.5898   |

| γ         | Old ρ_U (%) | Old ρ_L (%) | New ρ_U (%) | New ρ_L (%) |
|-----------|-------------|-------------|-------------|-------------|
| 0.1       | 0.2         | 0.4         | 0.0         | 0.0         |
| 0.05      | 0.2         | 0.6         | 0.0         | 0.0         |
| 0.01      | 0.4         | 1.0         | 0.0         | 0.0         |
| 0.000000002 | 1.4       | 5.9         | 0.2         | 0.1         |
| 5.421e–20 | 2.9         | 19.1        | 0.4         | 0.7         |

arising from one of quadratic formulae presented in this paper, we define the relative error ρ_U as

\[ ρ_U := \frac{1 + \tilde{δ}_U}{1 + δ_U} - 1. \]

Similarly, for δ_L an exact solution to one of the transcendental equations (2), (15) and \( \tilde{δ}_L \) an approximate solution to these arising from one of quadratic formulae presented in this paper, we define the relative error ρ_L as

\[ ρ_L := 1 - \frac{1 - \tilde{δ}_L}{1 - δ_L}. \]

As we expected, our new approximation is closer to the exact Chernoff bound, particularly for larger δ which correspond to smaller γ or smaller μ. For the smallest γ value, we observe that the upper bounds for the number of successful trials for the different methods are 347 (exact Chernoff), 357 (old approximation), and 348 (new approximation) and that the lower bound is 83, 67, and 83 respectively. Looking at the relative error table, we see that although the new estimate does outperform the old estimate at all levels, the proportional effect is not great until we reach the sort of tail probabilities associated with security. With these extremely small tail probabilities, there is a noticeable benefit to the new approximation, especially if further calculation is done with the results.

Turning now to our regression estimates, we assume that we run 1,000,000 Bernoulli trials with unknown probability and that we observe \( \hat{μ} = 212 \) successes. Again we choose confidence levels \( 1 - γ \) with γ values 0.1, 0.05, 0.01, 2.0e–9, 5.421e–20. For

2 We again observe that the Chernoff bounds are computed using only the number of observed successes and do not directly depend on the number of trials.
Table 2 Comparison of confidence bounds

| $\gamma$ | Exact $\delta_U$ | Exact $\delta_L$ | Old $\delta_U$ | Old $\delta_L$ | New $\delta_U$ | New $\delta_L$ |
|----------|-----------------|-----------------|----------------|----------------|----------------|----------------|
| 0.1      | 0.1547          | 0.1402          | 0.1586         | 0.1474         | 0.1548         | 0.1402         |
| 0.05     | 0.1777          | 0.1588          | 0.1828         | 0.1681         | 0.1778         | 0.1588         |
| 0.01     | 0.2232          | 0.1942          | 0.2313         | 0.2084         | 0.2234         | 0.1942         |
| 0.000000002 | 0.4998      | 0.3741          | 0.5393         | 0.4347         | 0.5022         | 0.3746         |
| 5.421e–20 | 0.7933        | 0.5156          | 0.8892         | 0.6469         | 0.8013         | 0.5176         |

| $\gamma$ | Old $\rho_U$ (%) | Old $\rho_L$ (%) | New $\rho_U$ (%) | New $\rho_L$ (%) |
|----------|------------------|------------------|------------------|------------------|
| 0.1      | 0.3              | 0.8              | 0.0              | 0.0              |
| 0.05     | 0.4              | 1.1              | 0.0              | 0.0              |
| 0.01     | 0.7              | 1.8              | 0.0              | 0.0              |
| 0.000000002 | 2.6        | 9.7              | 0.2              | 0.1              |
| 5.421e–20 | 5.3              | 27.1             | 0.4              | 0.4              |

Each $\gamma$ we compute $\delta_U$ and $\delta_L$ using the exact transcendental Chernoff formulae, the bounds analogous to existing quadratic formulae and the new quadratic formulae of this paper. Solutions are given to four significant figures in Table 2.

We again see very close values. Again we note that for the smallest $\gamma$ value the upper bound for a confident $\mu$ value is roughly 0.0003802 in the exact Chernoff case and 0.0003819 using our approximation and that the lower bound is roughly 0.0001027 in the exact Chernoff bound and 0.0001203 using our approximation. Looking at the relative error table, we see that again the benefit of the new estimates is not great until we reach extremely high levels of confidence, but at such levels the improvement is more pronounced than the predictive case.

Visualising the advantages of the new method when the $\delta$ values depend on both $\gamma$ (the size of the tail probability/exception to confidence) and $\mu$ (the expected number of observations resp. $\hat{\mu}$ the actual number of observations) is inconvenient. If however we treat $\delta$ as a function of $\beta = (\log \gamma) / \mu$ (resp. $(\log \gamma) / \hat{\mu}$) per the theorem statements the behaviour of the estimates are easy to compare graphically.

In Figs. 1, 2, 3, and 4, we focus on ranges where $0 < \delta < 1$ for all three bounds. The transcendental Chernoff bound solved numerically to 4 significant figures is shown in solid black, the old quadratic approximation in dotted red and the new quadratic approximation in dashed blue. For all of the cases, we see that as $\beta$ tends to zero (which could arise from examining longer tails or having more expected observations), there is no great separation. At $\beta = -0.05$ (which might represent cases close to $(\mu, \gamma) = (60, 0.05), (\mu, \gamma) = (92, 0.01)$ or $(\mu, \gamma) = (400, 0.000000002)$) in all cases, there is already a discernible case for the new estimate. At $\beta = -0.25$ (which might represent cases close to $(\mu, \gamma) = (92, 10^{-10}), (\mu, \gamma) = (177, 2^{-64})$ or $(\mu, \gamma) = (396, 10^{-43})$), the new estimate is still very close to the transcendental value and the old estimate is palpably inferior. The effect is more pronounced for lower bounds and regression estimates than upper bounds and predictive estimates.
Efficient computation of tight approximations to Chernoff…

Fig. 1 $\beta$ versus $\delta$ for predictive upper bound (–, exact bound; red dotted lines, old bound; blue dashed lines, new bound)

Fig. 2 $\beta$ versus $\delta$ for predictive lower bound (–, exact bound; red dotted lines, old bound; blue dashed lines, new bound)

5 Computational efficiency

A natural question is whether the improved estimates come with an increased computational burden. To compare the computational cost of the new and old methods and also the cost of directly solving the exact transcendental equations, all of these processes were implemented in Python 3.9.2 in the IDLE environment running on an HP Z2 Tower G5 Workstation (Intel i7 -107000 CPU @ 3.80GHz). The transcendental equa-
tion was solved using a naïve binary search algorithm run on the interval (0, 1) looking for an answer accurate to 4 significant figures. All of the processes took $\beta = \log \gamma / \mu$ as an input and a small range of $\beta$ values was tested. Each process was repeatedly computed 10,000 times and the time taken for all 10,000 computations in seconds was recorded using Python’s `time.perf_counter()` method. These measurements were then repeated 100 times to give and mean (and standard deviation) time to run
Table 3 Comparison of computation times (predictive, 10,000 computations, times given in seconds)

| $\beta$   | Bound  | Search: mean | S.d. | Old: mean | S.d. | New: mean | S.d. |
|-----------|--------|--------------|------|-----------|------|-----------|------|
| -0.001467| Lower  | 8.885e-2     | 7.840e-4 | 1.316e-3  | 1.381e-4 | 2.000e-3  | 2.287e-4 |
| -0.001467| Upper  | 8.608e-2     | 7.393e-4 | 1.830e-3  | 4.765e-5 | 1.763e-3  | 6.492e-5 |
| -0.01467 | Lower  | 7.724e-2     | 6.647e-4 | 1.310e-3  | 5.721e-5 | 1.977e-3  | 4.426e-5 |
| -0.01467 | Upper  | 7.510e-2     | 5.435e-4 | 1.845e-3  | 9.819e-5 | 1.748e-3  | 5.135e-5 |
| -0.1467  | Lower  | 7.274e-2     | 6.178e-4 | 1.312e-3  | 8.247e-5 | 1.985e-3  | 6.427e-5 |
| -0.1467  | Upper  | 7.030e-2     | 8.301e-4 | 1.836e-3  | 6.695e-5 | 1.747e-3  | 4.408e-5 |

Table 4 Comparison of computation times (regressive, 10,000 computations, times given in seconds)

| $\beta$   | Bound  | Search: mean | S.d. | Old: mean | S.d. | New: mean | S.d. |
|-----------|--------|--------------|------|-----------|------|-----------|------|
| -0.001467| Lower  | 7.912e-2     | 7.397e-4 | 1.266e-3  | 3.605e-5 | 2.131e-3  | 1.554e-4 |
| -0.001467| Upper  | 8.501e-2     | 8.872e-4 | 1.816e-3  | 6.933e-5 | 1.774e-3  | 3.881e-5 |
| -0.01467 | Lower  | 6.879e-2     | 8.515e-4 | 1.275e-3  | 9.612e-5 | 2.129e-3  | 1.629e-4 |
| -0.01467 | Upper  | 7.588e-2     | 1.540e-3 | 1.856e-3  | 1.547e-4 | 1.791e-3  | 1.852e-4 |
| -0.1467  | Lower  | 6.525e-2     | 6.556e-4 | 1.264e-3  | 3.820e-5 | 2.122e-3  | 8.064e-5 |
| -0.1467  | Upper  | 6.641e-2     | 1.274e-3 | 1.815e-3  | 4.511e-5 | 1.812e-3  | 1.320e-4 |

Each computation 10,000 times in seconds. The timings were performed for both the predictive and regressive equations.

It is clear that both the old and new methods significantly outperform the binary search solution by at least an order of magnitude. The old lower bounds can be computed marginally quicker as the original equations for $\delta_L$ do not involve the additional additions, subtractions and divisions of a general quadratic equation. The timings for the new lower bound are of the same order of magnitude however. For the upper bounds, the new process does seem to consistently outperform the old, but the reasons for the improvement are unclear and the speed-up is very marginal indeed. All of the methods behave similarly for different $\beta$ inputs and the run times seem stable. Both the old and new methods do seem well-suited to efficiently investigating large parameter spaces if necessary (Tables 3, 4).

6 Higher degree approximation

Eager readers will be aware that cubic and quartic equation also admit closed form solutions. For still greater accuracy, higher degree Padé approximations could be used. We note the following Padé approximations that could be used to this end. In general, an approximation with numerator of degree $d_0$ and denominator of degree $d_1$ will admit an error term of size $O(x^{d_0+d_1+1})$ and in cases where the quadratic approximations of this paper give values of $\delta$ close to 1, such higher degree approximations may be appropriate. These approximations are easy to compute via the extended Euclidean algorithm.
for polynomials as implemented in the `sagemath PowerSeries pade(m,n)` method for example.

### 6.1 Predictive upper bound

\[
x - (1 + x) \log(1 + x) < \frac{-15x^2 - 7x^3}{30 + 24x + 3x^2}
\]

\[
x - (1 + x) \log(1 + x) < \frac{-210x^2 - 200x^3 - 35x^4}{420 + 540x + 180x^2 + 12x^3}
\]

### 6.2 Predictive lower bound

\[
-x - (1 - x) \log(1 - x) < \frac{-210x^2 + 125x^3}{420 - 390x + 60x^2 + 3x^3}
\]

\[
-x - (1 - x) \log(1 - x) < \frac{7350x^2 - 8260x^3 + 1975x^4}{-14700 + 21420x - 8640x^2 + 780x^3 + 18x^4}
\]

### 6.3 Regressive upper bound

\[
-x + \log(1 + x) < \frac{240x^2 + 155x^3}{-480 - 630x - 180x^2 + 3x^3}
\]

\[
-x + \log(1 + x) < \frac{-210x^2 - 220x^3 - 45x^4}{420 + 720x + 360x^2 + 48x^3}
\]

### 6.4 Predictive lower bound

\[
x + \log(1 - x) < \frac{-240x^2 + 155x^3}{480 = 630x + 180x^2 + x^3}
\]

\[
x + \log(1 - x) < \frac{3150x^2 - 3780x^3 + 985x^4}{-6300 + 11760 - 6660x^2 + 1080x^3 + 6x^4}
\]

**Acknowledgements**  The author is very grateful to Sophie Stevens for discussions on this work and also to an anonymous referee for several very useful suggestions for its improvement.

**References**

Bernstein S (1924) On a modification of Chebyshev’s inequality and of the error formula of Laplace. Uchenye Zapiski Nauch-Issled Kaf Ukraine Sect Math 1:38–48

Bindel N, Schanck JM (2020) Decryption failure is more likely after success. PQCrypto 2020 (LNCS 12100), pp 206–225

Dubhashi D, Panconesi A (2009) Concentration of measure for the analysis of randomized algorithms, 1st edn. Cambridge University Press, Cambridge
Mitzenmacher M, Upfal E (2017) Probability and computing: randomized algorithms and probabilistic analysis, 2nd edn. Cambridge University Press, Cambridge
Pirandola S (2021) Limits and security of free-space quantum communications. Phys Rev Res 3:013279
Zhang Z, Zhao Q, Razavi M, Ma X (2017) Improved key rate bounds for practical decoy-state quantum key distribution systems. Phys Rev A 95:012333

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.