Improved Probabilistic Pigeonhole Principle

Vadim Ponomarenko
Charles Sotelo

July 24, 2026

1 Introduction

The Pigeonhole Principle, famously attributed to Johann Peter Gustav Lejeune Dirichlet in 1834, states that placing T pigeons into m holes (where $T > m$) guarantees that at least one hole contains multiple pigeons. A probabilistic relaxation of this deterministic concept naturally intersects with the Birthday Problem, first appearing in print in 1939 (in [2]) by Richard von Mises, which calculates the probability of shared birthdays within a group. Merging these classical concepts yields the Probabilistic Pigeonhole Principle.

For any positive integer m and for any real $p \in (0, 1)$, we define $T = T(m, p)$ as the minimum integer such that placing at least T pigeons uniformly at random into m pigeonholes will give probability at least p that at least two pigeons will share a pigeonhole. Note that $m^{\underline{T}}/m^T < 1 - p \leq m^{\underline{T-1}}/m^{T-1}$, where $m^{\underline{T}} = m(m-1)\cdots(m-T+1)$ denotes the falling factorial.

While a closed form for T is not known, Bhattacharya (in [1]) recently established a closed-form upper bound. To streamline our notation, we define the radical term $R = R(m, p)$ as

$$R = \sqrt{2m \log\left(\frac{1}{1-p}\right) + \frac{1}{4}}.$$

Theorem 1 ([1])**.** *Given $m \in \mathbb{N}$ and $p \in (0, 1)$, $\lceil A(m, p)\rceil \geq T(m, p)$, where $A(m, p) = \frac{1}{2} + R$.*

Our main result, Theorem 2, improves this bound $A(m, p)$ to the bound $B(m, p)$, defined as

$$B(m, p) = \frac{1}{2} + \sqrt{\frac{R^2 - 1/4}{1 + R/(3m)} + \frac{1}{4}}.$$

We also analyze the sharpness of B. In Theorem 3, we prove that for $p \in (0, 0.95)$ and all $m \geq 2$, as well as for fixed $p \in (0, 1)$ and sufficiently large m, the gap $\lceil B\rceil - T \leq 1$. Thus, our estimate is remarkably sharp, whereas $\lceil A\rceil - T$ can grow arbitrarily large. Finally, in Section 4, we compute the improvement $A - B$.

2 Main Theorem

We first prove a technical lemma needed for our main theorem.

Lemma 1. *Let $m \in \mathbb{N}$, $p \in (0, 1)$, and $n \in \mathbb{Z}$ satisfying $n \geq B(m, p)$. Then*

$$\frac{n(n-1)}{2m}\left(1 + \frac{R}{3m}\right) \geq \log\left(\frac{1}{1-p}\right); \textit{ and}$$

$$\frac{R}{3m} \le \frac{2n-1}{6m} + \frac{n(n-1)}{6m^2}.$$

Proof. We start with our hypothesis $n \ge B(m,p)$, subtract $\frac{1}{2}$ from both sides, square the result, and subtract $\frac{1}{4}$ from both sides to get

$$n(n-1) \ge \frac{R^2 - 1/4}{1 + R/(3m)}. \tag{1}$$

Multiplying by the denominator gives

$$n(n-1)\left(1 + \frac{R}{3m}\right) \ge R^2 - \frac{1}{4} = 2m \log\left(\frac{1}{1-p}\right),$$

which rearranges to the first inequality.

To prove the second inequality, we now assume, for the sake of contradiction, that $\frac{R}{3m} > \frac{2n-1}{6m} + \frac{n(n-1)}{6m^2}$. Multiplying by $3m$ yields $R > n - \frac{1}{2} + \frac{n(n-1)}{2m}$. Subtracting $n - \frac{1}{2}$ from both sides and multiplying by $R + n - \frac{1}{2}$ gives:

$$R^2 - \left(n - \frac{1}{2}\right)^2 > \frac{n(n-1)}{2m}\left(R + n - \frac{1}{2}\right) > \frac{n(n-1)}{2m} R \tag{2}$$

Multiplying (1) by its denominator and rearranging, we get:

$$n(n-1)\frac{R}{3m} \ge R^2 - \frac{1}{4} - n(n-1) = R^2 - \left(n - \frac{1}{2}\right)^2 \tag{3}$$

Combining (2) with (3) gives $\frac{n(n-1)}{3m}R > \frac{n(n-1)}{2m}R$, and hence $\frac{1}{3} > \frac{1}{2}$, a contradiction. ∎

We are now ready for our main result.

Theorem 2. *Given $m \in \mathbb{N}$ and $p \in (0,1)$, $\lceil B(m,p) \rceil \ge T(m,p)$. Equivalently, if $n \ge B(m,p)$ pigeons are placed uniformly at random into m pigeonholes, the probability of at least one collision is greater than p.*

Proof. The probability that all n pigeons land in distinct pigeonholes is $m^{\underline{n}}/m^n$. It suffices to show this is less than $1-p$, or, equivalently, to show that $\log\left(\frac{1}{1-p}\right) < -\log\left(\frac{m^{\underline{n}}}{m^n}\right)$. Combining the two parts of Lemma 1, we get

$$\log\left(\frac{1}{1-p}\right) \le \frac{n(n-1)}{2m}\left(1 + \frac{R}{3m}\right) \le \frac{n(n-1)}{2m}\left(1 + \frac{2n-1}{6m} + \frac{n(n-1)}{6m^2}\right).$$

Applying Faulhaber's formulas for the sums of first, second, and third powers, gives

$$\frac{n(n-1)}{2m}\left(1 + \frac{2n-1}{6m} + \frac{n(n-1)}{6m^2}\right) = \sum_{j=1}^{n-1}\left(\frac{j}{m} + \frac{j^2}{2m^2} + \frac{j^3}{3m^3}\right).$$

Now, each summand $\left(\frac{j}{m} + \frac{j^2}{2m^2} + \frac{j^3}{3m^3}\right)$ is a truncated Taylor series for, and hence less than, $-\log\left(1 - \frac{j}{m}\right)$. Combining, we get the desired

$$\log\left(\frac{1}{1-p}\right) < \sum_{j=1}^{n-1} -\log\left(1 - \frac{j}{m}\right) = -\log\left(\frac{m^{\underline{n}}}{m^n}\right).$$

∎

3 Comparison to the ground truth

Throughout this section, let $m \in \mathbb{Z}$ with $m \geq 2$, $p \in (0,1)$, and set $r = \lceil B(m,p) \rceil - 2$. We need two lemmas; the first is technical while the second is our workhorse (the main result but with different hypotheses).

Lemma 2. *Let* $0 \leq x \leq \frac{3}{4}$. *Then* $-\log(1-x) \leq x + \frac{x^2}{2} + x^3$.

Proof. The Taylor series expansion gives $-\log(1-x) = x + \frac{x^2}{2} + \sum_{i=3}^{\infty} \frac{x^i}{i}$. We calculate as follows.

$$\sum_{i=3}^{\infty} \frac{x^i}{i} = x^3 \left(\frac{1}{3} + \frac{x}{4} + \sum_{i=5}^{\infty} \frac{x^{i-3}}{i} \right) \leq x^3 \left(\frac{1}{3} + \frac{x}{4} + \sum_{i=5}^{\infty} \frac{x^{i-3}}{5} \right)$$

$$= x^3 \left(\frac{1}{3} + \frac{x}{4} + \frac{x^2}{5(1-x)} \right) \leq x^3 \left(\frac{1}{3} + \frac{(\frac{3}{4})}{4} + \frac{(\frac{3}{4})^2}{5(1-(\frac{3}{4}))} \right) = \frac{233}{240} x^3 \leq x^3$$

■

Lemma 3. *Let* $m, p, r, B(m,p), T(m,p)$ *be as above. Suppose that* $\frac{r-1}{m} \leq \frac{3}{4}$ *and* $r(r-1)^2 < 2m(2m+r)$. *Then* $\lceil B(m,p) \rceil \leq T(m,p) + 1$.

Proof. If $r \leq 1$, the result is immediate. Hence, assume $r \geq 2$. Multiplying both sides of $r(r-1)^2 < 2m(2m+r)$ by $\frac{r}{4m^3}$, we obtain

$$\frac{r^2(r-1)^2}{4m^3} < \frac{2mr + r^2}{2m^2} = \frac{r}{m} + \frac{r^2}{2m^2}.$$

Adding $\sum_{j=1}^{r-1} \frac{j}{m} + \sum_{j=1}^{r-1} \frac{j^2}{2m^2}$ to both sides yields

$$\sum_{j=1}^{r-1} \frac{j}{m} + \sum_{j=1}^{r-1} \frac{j^2}{2m^2} + \frac{r^2(r-1)^2}{4m^3} < \sum_{j=1}^{r} \frac{j}{m} + \sum_{j=1}^{r} \frac{j^2}{2m^2}.$$

Applying Faulhaber's formulas to both sides, we have

$$\sum_{j=1}^{r-1} \frac{j}{m} + \sum_{j=1}^{r-1} \frac{j^2}{2m^2} + \sum_{j=1}^{r-1} \frac{j^3}{m^3} < \frac{r(r+1)}{2m} \left(1 + \frac{r + \frac{1}{2}}{3m} \right). \tag{4}$$

Since $\frac{r-1}{m} \leq \frac{3}{4}$, we can apply Lemma 2 to obtain

$$\sum_{j=1}^{r-1} -\log\left(1 - \frac{j}{m}\right) \leq \sum_{j=1}^{r-1} \frac{j}{m} + \sum_{j=1}^{r-1} \frac{j^2}{2m^2} + \sum_{j=1}^{r-1} \frac{j^3}{m^3} \tag{5}$$

Combining (4) and (5) gives

$$\sum_{j=1}^{r-1} -\log\left(1 - \frac{j}{m}\right) < \frac{r(r+1)}{2m} \left(1 + \frac{r + \frac{1}{2}}{3m} \right).$$

From $\lceil x \rceil < x + 1$ we have $r < B - 1$. Adding $\frac{1}{2}$ to both sides we get

$$r + \frac{1}{2} < B - \frac{1}{2} < A - \frac{1}{2} = R. \tag{6}$$

Returning to $r + \frac{1}{2} < B - \frac{1}{2}$, squaring, and subtracting $\frac{1}{4}$, we obtain

$$r(r+1) < \frac{R^2 - 1/4}{1 + R/(3m)} < R^2 - \frac{1}{4}. \tag{7}$$

Combining equations (6) and (7) yields

$$\frac{r(r+1)}{2m}\left(1 + \frac{r + \frac{1}{2}}{3m}\right) < \frac{1}{2m}\left(\frac{R^2 - 1/4}{1 + R/(3m)}\right)\left(1 + \frac{R}{3m}\right) = \log\left(\frac{1}{1-p}\right).$$

Therefore, we have

$$\log\left(\frac{1}{1-p}\right) > \sum_{j=1}^{r-1} - \log\left(1 - \frac{j}{m}\right) = -\log\left(\frac{m^{\underline{r}}}{m^r}\right).$$

Hence, r pigeons are insufficient to guarantee a collision, so $T(m,p) > r$. Since both are integers, $T(m,p) \geq r + 1 = \lceil B(m,p) \rceil - 1$, and the desired result follows. ∎

We are now ready to give broad conditions under which our upper bound is within one of ground truth.

Theorem 3. *Let $m \geq 2$ and $p \in (0,1)$. Let $T(m,p)$ and $B(m,p)$ be as above (the ground truth and our bound, respectively), and $r = \lceil B(m,p) \rceil - 2$. Suppose that (i) $p < 0.95$; or (ii) $m \geq \frac{1}{2}\log^3\left(\frac{1}{1-p}\right)$. Then*

$$\lceil B(m,p) \rceil \leq T(m,p) + 1.$$

Proof. Let $r = \lceil B \rceil - 2$. If $r \leq 1$, the result is immediate, so we assume $r \geq 2$. We will start by adding $\frac{1}{2}$ to both sides of $r < B - 1$, squaring, and subtracting $\frac{1}{4}$, to get the useful inequality

$$r(r+1) < \frac{2m \log\left(\frac{1}{1-p}\right)}{1 + R/(3m)} < 2m \log\left(\frac{1}{1-p}\right). \tag{8}$$

We now proceed, for each of (i) and (ii), to apply Lemma 3, for which we need to verify two inequalities.

First we take (i). We combine $p < 0.95$ with (8) to get $r(r+1) < 2m \log\left(\frac{1}{1-p}\right) < 2m \log\left(\frac{1}{1-0.95}\right) < 6m$. This forces $r \leq m$ (otherwise $r(r+1) \geq (m+1)(m+2) \geq 6m$, a contradiction). For $r \leq 5$, we verify arithmetically that $\frac{r-1}{m} \leq \frac{3}{4}$. For $r \geq 6$, because $(r-1)(r-6)+2 \geq 0$ rearranges to $8(r-1) \leq r(r+1) < 6m$, which gives $\frac{r-1}{m} < \frac{3}{4}$. Finally, since $(r-1)^2 < 6m$ and $r \leq m$, we have $r(r-1)^2 < 6mr = 2m(2r+r) \leq 2m(2m+r)$.

Now suppose that $p \geq 0.95$ and (ii) holds. We take $p \geq 0.95 > 1 - e^{-8/3}$, which rearranges to $\frac{1}{1-p} > e^{8/3}$, hence $\log\frac{1}{1-p} > \frac{8}{3}$. We square both sides, multiply by $\log\frac{1}{1-p}$, and rearrange, to get $\frac{1}{2}\log^3\left(\frac{1}{1-p}\right) > \frac{32}{9}\log\left(\frac{1}{1-p}\right)$. Hence (ii) also implies $m \geq \frac{32}{9}\log\left(\frac{1}{1-p}\right)$; we need both inequalities.

We apply (8) to get $r^2 < r(r+1) < 2m\log\left(\frac{1}{1-p}\right)$, and hence $\frac{r}{m} < \sqrt{\frac{2\log\left(\frac{1}{1-p}\right)}{m}}$. We now rearrange $m \geq \frac{32}{9}\log\left(\frac{1}{1-p}\right)$ to get $\sqrt{\frac{2\log\left(\frac{1}{1-p}\right)}{m}} \leq \frac{3}{4}$. Hence $\frac{r-1}{m} < \frac{r}{m} < \frac{3}{4}$. Returning to $r^2 < 2m\log\left(\frac{1}{1-p}\right)$, we raise both sides to the 3/2 power, getting $r^3 < \left(2m\log\left(\frac{1}{1-p}\right)\right)^{3/2}$. We now rearrange $m \geq \frac{1}{2}\left(\log\left(\frac{1}{1-p}\right)\right)^3$ to get $4m^2 \geq \left(2m\log\left(\frac{1}{1-p}\right)\right)^{3/2}$. Hence $r^3 < 4m^2$, and we conclude $r(r-1)^2 < r^3 < 4m^2 < 2m(2m+r)$. ∎

4 Comparison to the Previous Bound

We now measure the improvement of our new bound against the bound in [1]. It shows that the gap $A - B$ is less than, but asymptotically equal to, $\frac{1}{3}\log\left(\frac{1}{1-p}\right)$.

Theorem 4. *Let $m \geq 2$, $p \in (0,1)$, and A, B, R be as above. Then*

$$\frac{1}{1+R/(3m)} < \frac{A-B}{\frac{1}{3}\log\left(\frac{1}{1-p}\right)} < 1.$$

Proof. We simplify, rationalize with the conjugate, and use $R^2 - \frac{1}{4} = 2m\log\left(\frac{1}{1-p}\right)$ to get

$$A - B = R - \sqrt{\frac{R^2-\frac{1}{4}}{1+R/(3m)} + \frac{1}{4}} = \frac{R^2 - \frac{R^2-1/4}{1+R/(3m)} - \frac{1}{4}}{R+B-\frac{1}{2}} = \frac{\left(2\log\left(\frac{1}{1-p}\right)\right)\left(\frac{R}{3+R/m}\right)}{R+B-\frac{1}{2}}$$

Dividing by $\frac{1}{3}R\log\left(\frac{1}{1-p}\right)$ and simplifying, we get the following.

$$\frac{A-B}{\frac{1}{3}\log\left(\frac{1}{1-p}\right)} = \left(\frac{1}{1+R/(3m)}\right)\left(\frac{2}{1+\sqrt{X}}\right), \text{ where } X = \frac{1+\frac{1}{12Rm}}{1+\frac{R}{3m}}. \tag{9}$$

Since $R > \frac{1}{2}$, we find $\frac{1}{12Rm} < \frac{R}{3m}$, and hence

$$\frac{1}{1+R/(3m)} < X < 1. \tag{10}$$

We complete the proof by using (9), combined with bounds on $\frac{2}{1+\sqrt{x}}$ using (10). Since $X < 1$, $1+\sqrt{X} < 2$, and hence $\frac{2}{1+\sqrt{X}} > 1$. This establishes the theorem's first inequality. For convenience, set $y = \sqrt{1+R/(3m)} > 1$. Since $X > \frac{1}{y^2}$, $\sqrt{X} > \frac{1}{y}$. Because $y > 1$ we get $y(y+1) > 2$, which rearranges to $1 + \frac{1}{y} > \frac{2}{y^2}$. Hence $1+\sqrt{X} > 1+\frac{1}{y} > \frac{2}{y^2} = \frac{2}{1+R/(3m)}$, which rearranges to $1+R/(3m) > \frac{2}{1+\sqrt{X}}$. This establishes the theorem's second inequality. ∎

Corollary 1. *Fix $p \in (0,1)$, and let A, B, R be as above. Then $\lim_{m\to\infty}(A-B) = \frac{1}{3}\log\left(\frac{1}{1-p}\right)$.*

Proof. Immediate from Theorem 4, since $R = O(m^{1/2})$. ∎

The new bound B provides its largest improvement over A when p is close to 1. To illustrate this, we recall a corollary from [1], which we improve by 1.

Corollary 2 ([1])**.** *If $3n+1$ or more pigeons are placed in n^2 pigeonholes randomly, then the probability that at least two pigeons are in the same hole is greater than 0.98889.*

Corollary 3. *If $3n$ or more pigeons are placed in n^2 pigeonholes randomly, then the probability that at least two pigeons are in the same hole is greater than 0.98889.*

Proof. Here $m = n^2$ and $p = 0.98889$. Since $\log(\frac{1}{1-p}) = 4.4999099675 < \frac{9}{2}$, we calculate

$$B(n^2,p) < \frac{1}{2} + \sqrt{\frac{9n^2}{1+1/n} + \frac{1}{4}} < \frac{1}{2} + \sqrt{9n^2 - 3n + \frac{1}{4}} = 3n.$$

∎

References

[1] Soumya Bhattacharya. The probabilistic pigeonhole principle. *Amer. Math. Monthly*, 130(7):678, 2023.

[2] R. Mises. Ueber Aufteilungs- und Besetzungs-Wahrscheinlichkeiten. *Acta [Trudy] Univ. Asiae Mediae. Ser. V-a.*, 1939(Fasc.):21, 1939.