

The Loot-Box Mean Problem: A Probabilistic Analysis

1 Problem Statement

In an online game, each loot box contains a random number of gold coins. The number of coins in each box is uniformly distributed over the set

$$\{1, 2, \dots, n\}. \quad (1)$$

A player wants to collect at least M gold coins.

We want to answer the following questions:

1. What is the expected number of loot boxes the player needs to open?
2. What is the distribution of the number of opened loot boxes?
3. What happens for small values $n = 1, 2, 3$?
4. What general formula can be obtained when $n \leq M$?
5. What happens in the special case $n = M$?
6. What happens when $n > M$?
7. What is the limit of the expectation when $n = M$ and $n \rightarrow \infty$?
8. How can the expectation be computed algorithmically?

2 The Probability Model

Let

$$X_1, X_2, X_3, \dots \quad (2)$$

be independent and identically distributed random variables.

The random variable X_i represents the number of gold coins in the i -th loot box. We assume

$$X_i \sim \text{Uniform}\{1, 2, \dots, n\}. \quad (3)$$

This means that

$$\mathbb{P}(X_i = j) = \frac{1}{n}, \quad j = 1, 2, \dots, n. \quad (4)$$

Define the partial sums

$$S_k = X_1 + X_2 + \dots + X_k, \quad k \geq 1, \quad (5)$$

and set

$$S_0 = 0. \quad (6)$$

Thus S_k is the total number of coins obtained after opening k loot boxes.

The player stops when the total number of coins is at least M . Hence we define

$$T = \min\{k \geq 1 : S_k \geq M\}. \quad (7)$$

The random variable T is the number of loot boxes opened.

3 Basic Probability Concepts

Definition 3.1 (Discrete random variable). A random variable is called discrete if its possible values form a finite or countable set.

In this problem, each X_i is discrete because it takes values in the finite set $\{1, 2, \dots, n\}$. The random variable T is also discrete.

Definition 3.2 (Expectation). If Y is a nonnegative integer-valued random variable, then its expectation is

$$\mathbb{E}[Y] = \sum_{k=0}^{\infty} k \mathbb{P}(Y = k). \quad (8)$$

The expectation is the long-run average value of a random variable.

In our problem, $T \leq M$, because every loot box contains at least one coin. Therefore, after M boxes, the player has at least M coins. Hence

$$\mathbb{P}(T \leq M) = 1. \quad (9)$$

So the expectation can be written as the finite sum

$$\mathbb{E}[T] = \sum_{t=1}^M t \mathbb{P}(T = t). \quad (10)$$

4 The Tail-Sum Formula

Theorem 4.1 (Tail-sum formula). *If Y is a positive integer-valued random variable, then*

$$\mathbb{E}[Y] = \sum_{k=0}^{\infty} \mathbb{P}(Y > k). \quad (11)$$

Proof. By the definition of expectation,

$$\mathbb{E}[Y] = \sum_{t=1}^{\infty} t \mathbb{P}(Y = t). \quad (12)$$

For every positive integer t ,

$$t = \sum_{k=0}^{t-1} 1. \quad (13)$$

Substituting (13) into (12), we get

$$\begin{aligned} \mathbb{E}[Y] &= \sum_{t=1}^{\infty} \left(\sum_{k=0}^{t-1} 1 \right) \mathbb{P}(Y = t) \\ &= \sum_{k=0}^{\infty} \sum_{t=k+1}^{\infty} \mathbb{P}(Y = t) \\ &= \sum_{k=0}^{\infty} \mathbb{P}(Y > k). \end{aligned} \quad (14)$$

This proves (11). □

Applying (11) to T , and using $T \leq M$, gives

$$\mathbb{E}[T] = \sum_{k=0}^{M-1} \mathbb{P}(T > k). \quad (15)$$

Now observe that

$$T > k \iff S_k < M. \quad (16)$$

Indeed, $T > k$ means that after k boxes the player has still not reached M coins. This is exactly the event $S_k < M$.

Therefore

$$\boxed{\mathbb{E}[T] = \sum_{k=0}^{M-1} \mathbb{P}(S_k < M).} \quad (17)$$

Formula (17) is one of the central formulas of the report.

5 Distribution of the Number of Loot Boxes

The distribution of T consists of the probabilities

$$\mathbb{P}(T = t), \quad t = 1, 2, \dots, M. \quad (18)$$

Theorem 5.1 (Distribution of T). *For $t = 1, 2, \dots, M$,*

$$\boxed{\mathbb{P}(T = t) = \mathbb{P}(S_{t-1} < M) - \mathbb{P}(S_t < M).} \quad (19)$$

Proof. The event $T = t$ means that the player has not reached M coins after $t - 1$ boxes, but has reached M coins after t boxes. Thus

$$\{T = t\} = \{S_{t-1} < M, S_t \geq M\}. \quad (20)$$

Since the sums S_k are increasing in k , the event $\{S_t < M\}$ is contained in $\{S_{t-1} < M\}$. Therefore

$$\{T = t\} = \{S_{t-1} < M\} \setminus \{S_t < M\}. \quad (21)$$

Taking probabilities gives (19). $\square$

6 First-Step Analysis and the Recurrence for E_m

We now derive a recursive formula for the expected number of additional boxes needed.

Definition 6.1. Let E_m be the expected number of additional loot boxes needed when the player still needs m more coins.

Thus the quantity we want is

$$\mathbb{E}[T] = E_M. \quad (22)$$

If $m \leq 0$, then the player already has enough coins. Hence no more boxes are needed:

$$E_m = 0, \quad m \leq 0. \quad (23)$$

For $m \geq 1$, the player opens one new box. That first box contributes one unit to the number of boxes opened. After opening it, there are several possible cases:

- If the box contains 1 coin, then the player still needs $m - 1$ coins.

- If the box contains 2 coins, then the player still needs $m - 2$ coins.
- More generally, if the box contains j coins, then the player still needs $m - j$ coins.

Since the loot box value is uniform on $\{1, 2, \dots, n\}$, the probability of each value j is $1/n$. Therefore the expected number of boxes after the first box is the average of the expectations E_{m-j} .

Theorem 6.2 (First-step recurrence). *For $m \geq 1$,*

$$E_m = 1 + \frac{1}{n} \sum_{j=1}^n E_{m-j}, \quad (24)$$

with boundary condition (23).

Proof. Let X be the number of coins in the next loot box. Then

$$\mathbb{P}(X = j) = \frac{1}{n}, \quad j = 1, 2, \dots, n. \quad (25)$$

If $X = j$, then after opening the box the remaining target is $m - j$. The number of additional boxes needed after that has expectation E_{m-j} .

The total number of boxes needed is therefore

$$1 + \text{number of boxes needed after the first box.} \quad (26)$$

By conditioning on the value of X , we obtain

$$\begin{aligned} E_m &= \sum_{j=1}^n \mathbb{P}(X = j) (1 + E_{m-j}) \\ &= \sum_{j=1}^n \frac{1}{n} (1 + E_{m-j}) \\ &= 1 + \frac{1}{n} \sum_{j=1}^n E_{m-j}. \end{aligned} \quad (27)$$

If $m - j \leq 0$, then $E_{m-j} = 0$ by (23). This covers the possibility that the first box already reaches or exceeds the target. $\square$

Remark 6.3. This method is called *first-step analysis*. It is common in probability problems where a process evolves step by step and the future after the first step has the same structure as the original problem.

7 A Short Theory of Nonhomogeneous Linear Recurrences

The formulas for $n = 2$ and $n = 3$ come from solving linear recurrences. We record the general principle.

Definition 7.1 (Linear recurrence with constant coefficients). A recurrence of the form

$$a_m = c_1 a_{m-1} + c_2 a_{m-2} + \dots + c_r a_{m-r} + g_m \quad (28)$$

is called a nonhomogeneous linear recurrence of order r with constant coefficients. The term g_m is called the nonhomogeneous term.

The associated homogeneous recurrence is

$$a_m = c_1 a_{m-1} + c_2 a_{m-2} + \cdots + c_r a_{m-r}. \quad (29)$$

Proposition 7.2 (Solution structure for nonhomogeneous recurrences). *Every solution of the nonhomogeneous recurrence (28) can be written as*

$$a_m = a_m^{(h)} + a_m^{(p)}, \quad (30)$$

where $a_m^{(h)}$ is a general solution of the homogeneous recurrence (29), and $a_m^{(p)}$ is one particular solution of the nonhomogeneous recurrence.

Proof. Suppose $a_m^{(p)}$ is one particular solution of (28). Let a_m be any other solution, and define

$$b_m = a_m - a_m^{(p)}. \quad (31)$$

Substituting $a_m = b_m + a_m^{(p)}$ into (28), and subtracting the equation satisfied by $a_m^{(p)}$, gives

$$b_m = c_1 b_{m-1} + c_2 b_{m-2} + \cdots + c_r b_{m-r}. \quad (32)$$

Thus b_m satisfies the homogeneous recurrence. Hence every solution is the sum of one particular solution and a homogeneous solution. $\square$

Proposition 7.3 (Characteristic equation method). *For the homogeneous recurrence*

$$a_m = c_1 a_{m-1} + c_2 a_{m-2} + \cdots + c_r a_{m-r}, \quad (33)$$

we try solutions of the form

$$a_m = \lambda^m. \quad (34)$$

Substitution gives the characteristic equation

$$\lambda^r - c_1 \lambda^{r-1} - c_2 \lambda^{r-2} - \cdots - c_r = 0. \quad (35)$$

If this equation has distinct roots $\lambda_1, \dots, \lambda_r$, then the homogeneous solution is

$$a_m^{(h)} = C_1 \lambda_1^m + \cdots + C_r \lambda_r^m. \quad (36)$$

Remark 7.4. In this report, the nonhomogeneous term is always the constant 1. Since the homogeneous recurrences for $n = 2$ and $n = 3$ have $\lambda = 1$ as a root, a constant particular solution would not work. Therefore we try a linear particular solution $a_m^{(p)} = cm$.

8 Small Values of n

8.1 Case $n = 1$

If $n = 1$, each loot box contains exactly one coin:

$$X_i = 1 \quad \text{with probability 1.} \quad (37)$$

Therefore

$$S_k = k. \quad (38)$$

The player reaches M coins exactly when $k = M$. Hence

$$T = M \quad \text{with probability 1.} \quad (39)$$

Therefore

$$\boxed{\mathbb{E}[T] = M.} \quad (40)$$

The distribution is

$$\boxed{\mathbb{P}(T = M) = 1.} \quad (41)$$

8.2 Case $n = 2$

For $n = 2$, recurrence (24) becomes

$$E_m = 1 + \frac{1}{2}E_{m-1} + \frac{1}{2}E_{m-2}, \quad m \geq 1, \quad (42)$$

with

$$E_0 = 0, \quad E_{-1} = 0. \quad (43)$$

Compute the first values.

For $m = 1$,

$$E_1 = 1 + \frac{1}{2}(E_0 + E_{-1}) = 1. \quad (44)$$

For $m = 2$,

$$E_2 = 1 + \frac{1}{2}(E_1 + E_0) = 1 + \frac{1}{2} = \frac{3}{2}. \quad (45)$$

For $m = 3$,

$$E_3 = 1 + \frac{1}{2}(E_2 + E_1) = 1 + \frac{1}{2}\left(\frac{3}{2} + 1\right) = \frac{9}{4}. \quad (46)$$

Proposition 8.1 (Explicit formula for $n = 2$). *For $n = 2$,*

$$E_M = \frac{2M}{3} + \frac{2}{9} - \frac{2}{9} \left(-\frac{1}{2}\right)^M. \quad (47)$$

Proof. Rewrite (42) as

$$E_m - \frac{1}{2}E_{m-1} - \frac{1}{2}E_{m-2} = 1. \quad (48)$$

The homogeneous recurrence is

$$E_m - \frac{1}{2}E_{m-1} - \frac{1}{2}E_{m-2} = 0. \quad (49)$$

Trying $E_m = r^m$, we get

$$r^2 - \frac{1}{2}r - \frac{1}{2} = 0. \quad (50)$$

Multiplying by 2, this becomes

$$2r^2 - r - 1 = 0. \quad (51)$$

Factoring,

$$2r^2 - r - 1 = (2r + 1)(r - 1). \quad (52)$$

Thus

$$r = 1, \quad r = -\frac{1}{2}. \quad (53)$$

Since the nonhomogeneous term in (48) is constant and $r = 1$ is a homogeneous root, we try

$$E_m^{(p)} = am. \quad (54)$$

Substituting $E_m = am$ into (48) gives

$$am - \frac{1}{2}a(m-1) - \frac{1}{2}a(m-2) = 1. \quad (55)$$

The left-hand side simplifies to

$$\frac{3}{2}a. \quad (56)$$

Therefore

$$\frac{3}{2}a = 1, \quad a = \frac{2}{3}. \quad (57)$$

The general solution is

$$E_m = \frac{2m}{3} + C + D \left(-\frac{1}{2}\right)^m. \quad (58)$$

Using $E_0 = 0$, we get

$$0 = C + D. \quad (59)$$

Using $E_1 = 1$, we get

$$1 = \frac{2}{3} + C - \frac{1}{2}D. \quad (60)$$

Solving (59) and (60) gives

$$C = \frac{2}{9}, \quad D = -\frac{2}{9}. \quad (61)$$

Substitution into (58) gives (47). $\square$

Remark 8.2. Formula (47) is correct. Substituting $M = 1, 2, 3$ gives 1, $3/2$, and $9/4$, agreeing with (44), (45), and (46).

8.3 Case $n = 3$

For $n = 3$, recurrence (24) becomes

$$E_m = 1 + \frac{1}{3}(E_{m-1} + E_{m-2} + E_{m-3}), \quad m \geq 1, \quad (62)$$

with

$$E_m = 0, \quad m \leq 0. \quad (63)$$

Compute the first values.

For $m = 1$,

$$E_1 = 1 + \frac{1}{3}(E_0 + E_{-1} + E_{-2}) = 1. \quad (64)$$

For $m = 2$,

$$E_2 = 1 + \frac{1}{3}(E_1 + E_0 + E_{-1}) = 1 + \frac{1}{3} = \frac{4}{3}. \quad (65)$$

For $m = 3$,

$$E_3 = 1 + \frac{1}{3}(E_2 + E_1 + E_0) = 1 + \frac{1}{3} \left(\frac{4}{3} + 1\right) = \frac{16}{9}. \quad (66)$$

For $m = 4$,

$$E_4 = 1 + \frac{1}{3}(E_3 + E_2 + E_1) = 1 + \frac{1}{3} \left(\frac{16}{9} + \frac{4}{3} + 1\right) = \frac{64}{27}. \quad (67)$$

Proposition 8.3 (Explicit formula for $n = 3$). *Define*

$$\lambda_+ = \frac{-1 + i\sqrt{2}}{3}, \quad \lambda_- = \frac{-1 - i\sqrt{2}}{3}. \quad (68)$$

Then

$$E_M = \frac{M}{2} + \frac{1}{3} + \left(-\frac{1}{6} - \frac{i\sqrt{2}}{24}\right) \lambda_+^M + \left(-\frac{1}{6} + \frac{i\sqrt{2}}{24}\right) \lambda_-^M. \quad (69)$$

Proof. Rewrite (62) as

$$E_m - \frac{1}{3}E_{m-1} - \frac{1}{3}E_{m-2} - \frac{1}{3}E_{m-3} = 1. \quad (70)$$

The homogeneous recurrence is

$$E_m - \frac{1}{3}E_{m-1} - \frac{1}{3}E_{m-2} - \frac{1}{3}E_{m-3} = 0. \quad (71)$$

Trying $E_m = r^m$, we get

$$r^3 - \frac{1}{3}r^2 - \frac{1}{3}r - \frac{1}{3} = 0. \quad (72)$$

Multiplying by 3,

$$3r^3 - r^2 - r - 1 = 0. \quad (73)$$

This factors as

$$3r^3 - r^2 - r - 1 = (r - 1)(3r^2 + 2r + 1). \quad (74)$$

Thus one root is $r = 1$. The other two roots solve

$$3r^2 + 2r + 1 = 0. \quad (75)$$

By the quadratic formula,

$$r = \frac{-2 \pm \sqrt{4 - 12}}{6} = \frac{-1 \pm i\sqrt{2}}{3}. \quad (76)$$

Hence the homogeneous solution has the form

$$C + A\lambda_+^m + B\lambda_-^m. \quad (77)$$

Since the nonhomogeneous term is constant and $r = 1$ is a homogeneous root, we try the particular solution

$$E_m^{(p)} = am. \quad (78)$$

Substituting $E_m = am$ into (70), we get

$$am - \frac{1}{3}a(m-1) - \frac{1}{3}a(m-2) - \frac{1}{3}a(m-3) = 1. \quad (79)$$

The left-hand side simplifies to

$$2a. \quad (80)$$

Thus

$$2a = 1, \quad a = \frac{1}{2}. \quad (81)$$

Therefore the general solution is

$$E_m = \frac{m}{2} + C + A\lambda_+^m + B\lambda_-^m. \quad (82)$$

Using $E_1 = 1$, $E_2 = 4/3$, and $E_3 = 16/9$, one obtains

$$C = \frac{1}{3}, \quad A = -\frac{1}{6} - \frac{i\sqrt{2}}{24}, \quad B = -\frac{1}{6} + \frac{i\sqrt{2}}{24}. \quad (83)$$

Substituting (83) into (82) gives (69). $\square$

Remark 8.4. Formula (69) is real-valued. Indeed, λ_- is the complex conjugate of λ_+ , and the coefficient of λ_-^M is the complex conjugate of the coefficient of λ_+^M . Thus the two complex terms add to a real number.

Remark 8.5. Formula (69) is correct. Substituting $M = 1, 2, 3, 4$ gives 1, 4/3, 16/9, and 64/27, agreeing with (64), (65), (66), and (67).

9 Counting Concepts

To obtain a closed formula for the distribution and expectation, we count the number of possible sequences of loot-box values.

9.1 Binomial coefficients

Definition 9.1 (Binomial coefficient). For integers $a \geq b \geq 0$, the binomial coefficient is

$$\binom{a}{b} = \frac{a!}{b!(a-b)!}. \quad (84)$$

It counts the number of ways to choose b objects from a objects, without caring about order.

Example 9.2. For example,

$$\binom{5}{2} = 10. \quad (85)$$

This is the number of two-element subsets of a five-element set.

Theorem 9.3 (Newton's binomial theorem). *For every integer $m \geq 0$,*

$$(1+x)^m = \sum_{k=0}^m \binom{m}{k} x^k. \quad (86)$$

Proof. When expanding $(1+x)^m$, each factor contributes either 1 or x . To obtain the term x^k , one must choose exactly k of the m factors to contribute x . This can be done in $\binom{m}{k}$ ways. Therefore the coefficient of x^k is $\binom{m}{k}$, proving (86). $\square$

9.2 Stars and bars

Proposition 9.4 (Stars and bars, equality form). *The number of nonnegative integer solutions of*

$$z_1 + z_2 + \cdots + z_k = r \quad (87)$$

is

$$\binom{r+k-1}{k-1}. \quad (88)$$

Proof. Imagine r identical stars arranged in a row. We want to split them into k groups, where the i -th group has size z_i . To create k groups, we place $k-1$ bars among the stars.

For example, with $r = 7$ and $k = 4$,

$$** \mid *** \parallel **$$

corresponds to

$$z_1 = 2, \quad z_2 = 3, \quad z_3 = 0, \quad z_4 = 2.$$

There are $r+k-1$ positions in total: r stars and $k-1$ bars. Choosing the positions of the $k-1$ bars determines the solution uniquely. Hence the number of solutions is

$$\binom{r+k-1}{k-1}. \quad (89)$$

Equivalently, choosing the positions of the r stars gives the same number. $\square$

Proposition 9.5 (Stars and bars, inequality form). *The number of nonnegative integer solutions of*

$$y_1 + y_2 + \cdots + y_k \leq r \quad (90)$$

is

$$\binom{r+k}{k}. \quad (91)$$

Proof. Introduce one extra nonnegative variable y_{k+1} and write

$$y_1 + y_2 + \cdots + y_k + y_{k+1} = r. \quad (92)$$

The variable y_{k+1} represents the unused part of the total r . Therefore solutions of (90) correspond exactly to nonnegative integer solutions of (92).

By the equality form of stars and bars, the number of solutions is

$$\binom{r + (k+1) - 1}{(k+1) - 1} = \binom{r+k}{k}. \quad (93)$$

□

10 The Inclusion-Exclusion Principle

Proposition 10.1 (Inclusion-exclusion principle). *Let $A_1, A_2, \dots, A_k$ be finite sets. Then*

$$\begin{aligned} |A_1 \cup A_2 \cup \cdots \cup A_k| &= \sum_i |A_i| - \sum_{i < j} |A_i \cap A_j| \\ &\quad + \sum_{i < j < \ell} |A_i \cap A_j \cap A_\ell| - \cdots \\ &\quad + (-1)^{k+1} |A_1 \cap A_2 \cap \cdots \cap A_k|. \end{aligned} \quad (94)$$

Proof. Fix an element x . Suppose x belongs to exactly q of the sets $A_1, \dots, A_k$. Then its total contribution to the right-hand side of (94) is

$$\binom{q}{1} - \binom{q}{2} + \binom{q}{3} - \cdots + (-1)^{q+1} \binom{q}{q}. \quad (95)$$

By the binomial theorem,

$$(1 - 1)^q = \sum_{j=0}^q (-1)^j \binom{q}{j} = 0. \quad (96)$$

Therefore

$$\binom{q}{1} - \binom{q}{2} + \cdots + (-1)^{q+1} \binom{q}{q} = 1. \quad (97)$$

So every element belonging to the union is counted exactly once. Elements not belonging to the union are counted zero times. Hence the formula is true. □

11 Counting $\mathbb{P}(S_k < M)$

We now count

$$\mathbb{P}(S_k < M). \quad (98)$$

The event $S_k < M$ means

$$X_1 + \cdots + X_k \leq M - 1. \quad (99)$$

Thus we must count the number of k -tuples

$$(x_1, x_2, \dots, x_k) \quad (100)$$

such that

$$1 \leq x_i \leq n \quad (101)$$

and

$$x_1 + x_2 + \dots + x_k \leq M - 1. \quad (102)$$

There are n^k total possible k -tuples, because each of the k boxes can take n possible values. Since all tuples are equally likely,

$$\mathbb{P}(S_k < M) = \frac{\#\{\text{valid } k\text{-tuples}\}}{n^k}. \quad (103)$$

We now transform the variables. Set

$$y_i = x_i - 1. \quad (104)$$

Then $1 \leq x_i \leq n$ becomes

$$0 \leq y_i \leq n - 1. \quad (105)$$

Also, (102) becomes

$$y_1 + \dots + y_k \leq M - 1 - k. \quad (106)$$

Set

$$r = M - 1 - k. \quad (107)$$

Thus we need to count the nonnegative integer solutions of

$$y_1 + \dots + y_k \leq r \quad (108)$$

subject to

$$y_i \leq n - 1 \quad \text{for all } i. \quad (109)$$

12 Applying Inclusion-Exclusion in Detail

First ignore the upper bounds (109). Let

$$U = \left\{ (y_1, \dots, y_k) \in \mathbb{Z}_{\geq 0}^k : y_1 + \dots + y_k \leq r \right\}. \quad (110)$$

By stars and bars,

$$|U| = \binom{r + k}{k}. \quad (111)$$

Using (107), this is

$$|U| = \binom{M - 1}{k}. \quad (112)$$

However, U includes solutions where some y_i may be too large. We must remove those.

For each $i = 1, 2, \dots, k$, define the bad set

$$A_i = \{ (y_1, \dots, y_k) \in U : y_i \geq n \}. \quad (113)$$

The valid solutions are exactly

$$U \setminus (A_1 \cup A_2 \cup \dots \cup A_k). \quad (114)$$

Therefore

$$\#\{\text{valid solutions}\} = |U| - |A_1 \cup A_2 \cup \dots \cup A_k|. \quad (115)$$

Now we use inclusion-exclusion to count the union of the bad sets.

Take a subset $J \subseteq \{1, 2, \dots, k\}$ with

$$|J| = j. \quad (116)$$

The intersection

$$\bigcap_{i \in J} A_i \quad (117)$$

is the set of solutions in U such that

$$y_i \geq n \quad \text{for every } i \in J. \quad (118)$$

For every $i \in J$, define

$$z_i = y_i - n. \quad (119)$$

Then $z_i \geq 0$. For variables not in J , leave them unchanged. This transformation subtracts n from each of the j selected variables, so the total allowed sum decreases by jn . Therefore the number of solutions in $\bigcap_{i \in J} A_i$ is the number of nonnegative solutions of

$$z_1 + \dots + z_k \leq r - jn. \quad (120)$$

By stars and bars, if $r - jn \geq 0$, this number is

$$\binom{r - jn + k}{k}. \quad (121)$$

Using $r = M - 1 - k$, this becomes

$$\binom{M - 1 - jn}{k}. \quad (122)$$

There are

$$\binom{k}{j} \quad (123)$$

ways to choose the subset J of size j .

Thus, by inclusion-exclusion, the number of valid solutions is

$$\sum_{j=0}^{\lfloor r/n \rfloor} (-1)^j \binom{k}{j} \binom{r - jn + k}{k}. \quad (124)$$

Substituting $r = M - 1 - k$, we obtain

$$\sum_{j=0}^{\lfloor \frac{M-1-k}{n} \rfloor} (-1)^j \binom{k}{j} \binom{M - 1 - jn}{k}. \quad (125)$$

Theorem 12.1 (Counting formula for $\mathbb{P}(S_k < M)$). *For $k = 0, 1, \dots, M - 1$,*

$$\mathbb{P}(S_k < M) = \frac{1}{n^k} \sum_{j=0}^{\lfloor \frac{M-1-k}{n} \rfloor} (-1)^j \binom{k}{j} \binom{M - 1 - jn}{k}. \quad (126)$$

For $k = 0$, we use the convention

$$\mathbb{P}(S_0 < M) = 1. \quad (127)$$

Proof. The proof is exactly the inclusion-exclusion argument above. Formula (125) gives the number of valid k -tuples. Dividing by the total number n^k of possible k -tuples gives (126). $\square$

13 General Expectation Formula for $n \leq M$

Combining the tail formula (17) with the probability formula (126), we get the exact expectation.

Theorem 13.1 (General expectation formula). *For $n \leq M$,*

$$\mathbb{E}[T] = \sum_{k=0}^{M-1} \frac{1}{n^k} \sum_{j=0}^{\lfloor \frac{M-1-k}{n} \rfloor} (-1)^j \binom{k}{j} \binom{M-1-jn}{k}. \quad (128)$$

Proof. From (17),

$$\mathbb{E}[T] = \sum_{k=0}^{M-1} \mathbb{P}(S_k < M). \quad (129)$$

Substituting (126) into (129) gives (128). $\square$

14 Checking the General Formula for $n = 1, 2, 3$

It is important to verify that the general formula (128) agrees with the small cases.

14.1 The case $n = 1$

If $n = 1$, every box contains exactly one coin. Thus $S_k = k$. Therefore

$$\mathbb{P}(S_k < M) = 1 \quad \text{for } k = 0, 1, \dots, M-1. \quad (130)$$

Hence (17) gives

$$\mathbb{E}[T] = \sum_{k=0}^{M-1} 1 = M, \quad (131)$$

which agrees with (40).

14.2 The case $n = 2$

For $n = 2$, formula (24) gives the recurrence (42). The first values computed from this recurrence are

$$E_1 = 1, \quad E_2 = \frac{3}{2}, \quad E_3 = \frac{9}{4}. \quad (132)$$

Formula (128) gives the same values because both methods compute the same expectation $\mathbb{E}[T]$. We can verify directly for $M = 3$. For $n = 2$ and $M = 3$,

$$\mathbb{E}[T] = \mathbb{P}(S_0 < 3) + \mathbb{P}(S_1 < 3) + \mathbb{P}(S_2 < 3). \quad (133)$$

Now

$$\mathbb{P}(S_0 < 3) = 1, \quad (134)$$

and

$$\mathbb{P}(S_1 < 3) = 1, \quad (135)$$

because one box contains either 1 or 2 coins.

For two boxes, the four equally likely outcomes are

$$(1, 1), \quad (1, 2), \quad (2, 1), \quad (2, 2). \quad (136)$$

Only $(1, 1)$ has sum less than 3. Hence

$$\mathbb{P}(S_2 < 3) = \frac{1}{4}. \quad (137)$$

Therefore

$$\mathbb{E}[T] = 1 + 1 + \frac{1}{4} = \frac{9}{4}, \quad (138)$$

which agrees with (46) and with (47) at $M = 3$.

14.3 The case $n = 3$

For $n = 3$, the recurrence gives

$$E_1 = 1, \quad E_2 = \frac{4}{3}, \quad E_3 = \frac{16}{9}. \quad (139)$$

Let us verify $M = 3$ using the tail formula. If $M = 3$, then

$$\mathbb{E}[T] = \mathbb{P}(S_0 < 3) + \mathbb{P}(S_1 < 3) + \mathbb{P}(S_2 < 3). \quad (140)$$

Clearly,

$$\mathbb{P}(S_0 < 3) = 1. \quad (141)$$

For one box,

$$S_1 < 3 \quad (142)$$

means the first box contains 1 or 2 coins. Thus

$$\mathbb{P}(S_1 < 3) = \frac{2}{3}. \quad (143)$$

For two boxes, each box can contain 1, 2, 3, so there are 9 equally likely pairs. The only pair with sum less than 3 is

$$(1, 1). \quad (144)$$

Hence

$$\mathbb{P}(S_2 < 3) = \frac{1}{9}. \quad (145)$$

Therefore

$$\mathbb{E}[T] = 1 + \frac{2}{3} + \frac{1}{9} = \frac{16}{9}, \quad (146)$$

which agrees with (66).

Thus the general expectation formula agrees with the explicit calculations for $n = 1, 2, 3$.

15 The Special Case $n = M$

Now suppose

$$n = M. \quad (147)$$

Before the player reaches the target M , any observed partial sum is strictly less than M . Thus, if

$$S_k < M, \quad (148)$$

then each individual value in that partial sum is automatically at most $M - 1$. Therefore the upper bound $X_i \leq M$ does not restrict the counting before the target is reached.

For $k = 0, 1, \dots, M - 1$, the number of positive integer k -tuples satisfying

$$x_1 + \dots + x_k \leq M - 1 \quad (149)$$

is

$$\binom{M-1}{k}. \quad (150)$$

Since there are M^k possible k -tuples in total,

$$\mathbb{P}(S_k < M) = \frac{1}{M^k} \binom{M-1}{k}. \quad (151)$$

Using (17),

$$\mathbb{E}[T] = \sum_{k=0}^{M-1} \binom{M-1}{k} \frac{1}{M^k}. \quad (152)$$

By Newton's binomial theorem (86),

$$\sum_{k=0}^{M-1} \binom{M-1}{k} \frac{1}{M^k} = \left(1 + \frac{1}{M}\right)^{M-1}. \quad (153)$$

Therefore:

Theorem 15.1 (Expectation when $n = M$). *If $n = M$, then*

$$\boxed{\mathbb{E}[T] = \left(1 + \frac{1}{M}\right)^{M-1}}. \quad (154)$$

16 The Case $n > M$

Now suppose

$$n > M. \quad (155)$$

If $S_k < M$, then every individual box value appearing in the partial sum is at most $M - 1$, and therefore automatically less than n . Hence the upper bound $X_i \leq n$ again does not restrict the counting before the target is reached.

Thus, just as in the case $n = M$, for $k = 0, 1, \dots, M - 1$,

$$\mathbb{P}(S_k < M) = \frac{1}{n^k} \binom{M-1}{k}. \quad (156)$$

Therefore

$$\mathbb{E}[T] = \sum_{k=0}^{M-1} \binom{M-1}{k} \frac{1}{n^k}. \quad (157)$$

By Newton's binomial theorem,

$$\sum_{k=0}^{M-1} \binom{M-1}{k} \frac{1}{n^k} = \left(1 + \frac{1}{n}\right)^{M-1}. \quad (158)$$

Theorem 16.1 (Expectation when $n \geq M$). *If $n \geq M$, then*

$$\boxed{\mathbb{E}[T] = \left(1 + \frac{1}{n}\right)^{M-1}}. \quad (159)$$

17 Why the Formulas Agree When $n = M$

One might worry that the general formula (128) and the formula (159) give different answers when $n = M$. They do not. They agree exactly.

Indeed, set $n = M$ in the inner sum of (128). For $k = 0, 1, \dots, M - 1$,

$$0 \leq M - 1 - k \leq M - 1. \quad (160)$$

Therefore

$$\left\lfloor \frac{M - 1 - k}{M} \right\rfloor = 0. \quad (161)$$

Thus the inner sum in (128) has only the term $j = 0$. Hence

$$\mathbb{P}(S_k < M) = \frac{1}{M^k} \binom{k}{0} \binom{M - 1}{k}. \quad (162)$$

Since $\binom{k}{0} = 1$, this becomes

$$\mathbb{P}(S_k < M) = \frac{1}{M^k} \binom{M - 1}{k}, \quad (163)$$

which is exactly (151). Therefore the general formula (128) reduces to

$$\mathbb{E}[T] = \sum_{k=0}^{M-1} \binom{M-1}{k} \frac{1}{M^k}. \quad (164)$$

By the binomial theorem, this is exactly

$$\left(1 + \frac{1}{M}\right)^{M-1}. \quad (165)$$

Therefore, there is no contradiction: the two formulas agree perfectly when $n = M$.

18 Distribution When $n \geq M$

When $n \geq M$, we have (156). Therefore, using (19),

$$\mathbb{P}(T = t) = \frac{\binom{M-1}{t-1}}{n^{t-1}} - \frac{\binom{M-1}{t}}{n^t}, \quad t = 1, 2, \dots, M. \quad (166)$$

Thus

$$\boxed{\mathbb{P}(T = t) = \frac{\binom{M-1}{t-1}}{n^{t-1}} - \frac{\binom{M-1}{t}}{n^t}.} \quad (167)$$

19 Limit When $n = M$ and $n \rightarrow \infty$

When $n = M$, equation (154) becomes

$$\mathbb{E}[T] = \left(1 + \frac{1}{n}\right)^{n-1}. \quad (168)$$

We compute

$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^{n-1}. \quad (169)$$

Recall the classical limit

$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = e. \quad (170)$$

Now

$$\left(1 + \frac{1}{n}\right)^{n-1} = \frac{\left(1 + \frac{1}{n}\right)^n}{1 + \frac{1}{n}}. \quad (171)$$

Taking limits in (171), and using (170), gives

$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^{n-1} = \frac{e}{1} = e. \quad (172)$$

Theorem 19.1 (Limiting expectation). *If $n = M$, then*

$$\mathbb{E}[T] = \left(1 + \frac{1}{n}\right)^{n-1}. \quad (173)$$

Consequently,

$$\boxed{\lim_{n \rightarrow \infty} \mathbb{E}[T] = e.} \quad (174)$$

20 Algorithms for Computing the Expectation

There are several practical ways to compute $\mathbb{E}[T]$.

20.1 Algorithm 1: Dynamic programming from the recurrence

The recurrence (24) gives a direct algorithm.

First set

$$E_0 = 0. \quad (175)$$

Also interpret

$$E_m = 0 \quad \text{for } m < 0. \quad (176)$$

Then compute $E_1, E_2, \dots, E_M$ using

$$E_m = 1 + \frac{1}{n} \sum_{j=1}^n E_{m-j}. \quad (177)$$

The final answer is

$$E_M. \quad (178)$$

This algorithm has complexity

$$O(Mn), \quad (179)$$

because for each m , one sums over n terms.

20.2 Algorithm 2: Sliding-window dynamic programming

We can improve the computation by defining

$$W_m = E_{m-1} + E_{m-2} + \dots + E_{m-n}. \quad (180)$$

Then

$$E_m = 1 + \frac{W_m}{n}. \quad (181)$$

When moving from m to $m + 1$, the window changes by adding E_m and removing E_{m-n} . Thus the update can be done in constant time. This gives an $O(M)$ algorithm.

20.3 Pseudocode

```

Input: n, M

Create an array E[0...M]
Set E[0] = 0

For m from 1 to M:
  sum = 0
  For j from 1 to n:
    If m-j >= 0:
      sum = sum + E[m-j]
  E[m] = 1 + sum/n

Output E[M]

```

This computes the expectation exactly up to numerical precision.

21 Further Research Ideas

Several extensions are natural.

1. Non-uniform loot boxes.

Instead of assuming a uniform distribution, we may assume

$$\mathbb{P}(X_i = j) = p_j, \quad j = 1, 2, \dots, n, \quad (182)$$

where

$$p_1 + p_2 + \dots + p_n = 1. \quad (183)$$

The recurrence becomes

$$E_m = 1 + \sum_{j=1}^n p_j E_{m-j}. \quad (184)$$

2. Variance of the number of boxes.

The expectation gives the average number of boxes, but one may also study

$$\text{Var}(T). \quad (185)$$

3. Probability of finishing within a fixed number of boxes.

One may compute

$$\mathbb{P}(T \leq r), \quad (186)$$

which is useful if a player has only r loot boxes available.

4. Overshoot analysis.

Usually, the player does not stop with exactly M coins, but with at least M . The overshoot is

$$S_T - M. \quad (187)$$

One may study its expectation and distribution.

5. Asymptotic analysis.

One can study the behaviour of $\mathbb{E}[T]$ as $M, n \rightarrow \infty$, especially when M/n tends to a fixed constant.

6. Cost models.

If each loot box costs money, one can study the expected cost of reaching M coins.

7. Adaptive reward systems.

Some games may change reward probabilities after many boxes. This leads to non-identically distributed random variables X_i .

22 Final Summary

Let

$$T = \min\{k \geq 1 : X_1 + \dots + X_k \geq M\}, \quad (188)$$

where

$$X_i \sim \text{Uniform}\{1, 2, \dots, n\}. \quad (189)$$

The distribution of T is

$$\mathbb{P}(T = t) = \mathbb{P}(S_{t-1} < M) - \mathbb{P}(S_t < M). \quad (190)$$

The expectation is

$$\mathbb{E}[T] = \sum_{k=0}^{M-1} \mathbb{P}(S_k < M). \quad (191)$$

For general $n \leq M$,

$$\mathbb{E}[T] = \sum_{k=0}^{M-1} \frac{1}{n^k} \sum_{j=0}^{\lfloor \frac{M-1-k}{n} \rfloor} (-1)^j \binom{k}{j} \binom{M-1-jn}{k}. \quad (192)$$

For $n \geq M$,

$$\mathbb{E}[T] = \left(1 + \frac{1}{n}\right)^{M-1}. \quad (193)$$

In particular, for $n = M$,

$$\mathbb{E}[T] = \left(1 + \frac{1}{n}\right)^{n-1}. \quad (194)$$

Finally,

$$\lim_{n \rightarrow \infty, n=M} \mathbb{E}[T] = e. \quad (195)$$

n	$\mathbb{E}(T)$	n	$\mathbb{E}(T)$
1	10.000000	11	2.188251
2	6.888672	12	2.055223
3	5.334756	13	1.948328
4	4.398823	14	1.860669
5	3.777380	15	1.787551
6	3.323694	16	1.725681
7	2.999510	17	1.672678
8	2.761508	18	1.626784
9	2.581175	19	1.586673
10	2.357948	20	1.551328

Table 1: Expected number of loot boxes $\mathbb{E}(T)$ needed to collect at least $M = 10$ coins, for $n = 1, 2, \dots, 20$. For $n \leq 10$, the values are computed from the general inclusion-exclusion formula; for $n \geq 10$, they agree with $\mathbb{E}(T) = (1 + \frac{1}{n})^9$.