HOW MANY DICE ROLLS DOES IT TAKE TO ACHIEVE ANY GOAL?

LUCY MARTINEZ

Abstract. HERE GOES THE ABSTRACT.

1. Introduction

A classical probability problem involves flipping a coin until a predetermined *goal* is reached. Suppose your goal is to keep flipping a fair coin until you obtain a certain number of heads (or tails). If your goal is to reach n heads, then this game can be modeled by the negative binomial distribution. The *negative binomial distribution* models how long it would take to get r successes in a succesion of n trials. If the probability of getting heads is p then the number of expected tosses to obtain n heads (or tails) is $\frac{n(1-p)}{p}$. For background on the negative binomial distribution, see [6, Chapter VI]. A classical problem that dates back to 1494 and laid the foundation for the theory of probability is called the *Problem of Points*. In modern times, we can state the problem as follows. Suppose that you toss a fair coin until you reach n heads or m tails. In 1954, the correspondence between Fermat and Pascal led to a solution to the probability of getting n heads vs. getting m tails. For more historical background on the Problem of Points, we refer the reader to an article by Edwards [5].

Jonathan Rosenberg studied the Problem of Points in a different context. In 2001, Rosenberg looked at the rules of Beit Hillel and Beit Shammai from the Talmud text on family planning [9]. We rephrase the rules of Beit Hillel and Beit Shammai in the context of coin tosses. Rosenberg studied the expected number of coin tosses for two scenarios. One scenario involves flipping a fair coin until you get one heads and one tails, while in the other scenario you stop once you get two heads. An interesting observation is that the expected number of coin tosses of each scenario is equal if and only if the probability of getting heads is the golden ratio. Later, Blatt, Freiberg, and Shikhman addressed a different question regarding the ratio of the expected number of heads to the expected number of tails when the goal is to reach n heads and m tails (in their notation $m = k$) [3]. In their work, they showed that the ratio of heads vs tails is $\frac{p}{1-p}$ where p represents the probability of getting heads.

More recently, Janson, Martinez, and Zeilberger studied the expected coin toss until you reach n heads or m tails [7]. Although this setup is the same as that of the Problem of points, Fermat and Pascal focused on the probability of getting n heads vs getting m tails, rather than on the duration of the game. Using probabilistic techniques, together with symbolic computation, Janson, Martinez, and Zeilberger derived probability generating functions, closed-form expressions, and asymptotic formulas for the moments of the stopping time to reach n heads or m tails. Other works that include reaching a predetermined goal include that of Noga Alon and Yaakov Malinovsky who studied the expected number of die rolls until the total sum of the outcomes is prime [1]. See also [2, 4, 8] for other variants and generalizations.

The goal of this paper is to extend previous generalizations of coin tosses to rolling a die to reach a predetermined *goal*. Consider rolling a loaded die with k faces and recording the number of appearances on each side. We study the expected number of dice rolls by developing recurrences that enable fast computation of expectations, variances, and higher moments. **Lucy: will finish introduction once everything is done**

2. Preliminaries

In this section, we introduce notation used throughout the paper. We let $\mathbb{N} \coloneqq \{1, 2, \ldots\}$ and $[k] \coloneqq \{1, 2, \ldots, k\}$. We use the following convention for the game:

- k denotes the number of faces on the loaded die, and $K \in \mathbb{N}$ is the cutoff representing the maximum number of die rolls allowed in a single game.

- A *game* consists of repeatedly rolling a loaded die until the goal is reached or the cutoff K is exceeded.
- For $1 \le i \le k$, let p_i be the probability corresponding that the die lands on side/face i, where $\sum_i p_i = 1$. We refer to this die as a *loaded die* and write $P = (p_1, p_2, \ldots, p_k)$ for the corresponding probability distribution.
- Let $s_0 = (0, 0, \ldots, 0) \in \mathbb{Z}^k_{\ge 0}$. For $t \ge 1$, let

$$s_t = (s_1, s_2, \ldots, s_k)$$

  denote the sequence of appearance counts after t rolls, where $s_{t,i}$ is the number of times the die landed on side i. For ease of notation, whenever we use s_i, we mean $s_{t,i}$.
- Let $x = (x_1, \ldots, x_k) \in \mathbb{Z}^k_{\ge 0}$ denote a vector of formal variables.
- Let $S = \{S_1, \ldots, S_\ell\}$ be a prescribed collection of subsets, where each

$$S_j = \{f_{j1}(x), \ldots, f_{jj'}(x)\},$$

  consists of linear functions of the form

$$f_{jm}(x) = a_1 x_1 + \cdots + a_k x_k - b,$$

  with $a_i \in \mathbb{Z}_{\ge 0}$ and $b \in \mathbb{N}$.
- The integer ℓ denotes the number of subsets in S, and j' denotes the number of functions in each subset S_j.

We are interested in repeatedly rolling a loaded k-sided die and tracking the resulting sequence $s_t = (s_1, s_2, \ldots, s_k)$. The game continues until there exists at least one index j, $1 \le j \le \ell$, such that evaluating the sequence s_t in every function of S_j yields nonnegative values. We define this formally next.

Definition 2.1 (Goal and stopping condition)**.** Let $s_0 = (0, 0, \ldots, 0) \in \mathbb{Z}^k_{\ge 0}$ be the initial state sequence. Let $(s_t)_{t \ge 0}$ be the sequence generated by repeatedly rolling a loaded k-sided die, where $s_t = (s_1, \ldots, s_k)$ records the number of times each face has appeared after t rolls. For each $j \in [\ell]$, define the feasible region or target region

$$A_j := \{x \in \mathbb{Z}^k_{\ge 0} : f(x) \ge 0 \text{ for all } f \in S_j\}, \quad A := \cup_{j=1}^{\ell} A_j.$$

We say that the game *reaches its goal at time t* if there exists at least one j, $1 \le j \le \ell$, such that $s_t \in A_j$. Equivalently, the game stops at the first time

$$\tau := \inf\{t : s_t \in \bigcup_{j=1}^{\ell} A_j\}.$$

When this occurs, we say that the sequence s_τ *enters* the feasible region or target region. We call s_τ the *terminal state.*

We stress that depending on the set S, different sequences may enter different target regions. The following example demonstrates this behavior.

Example 1. Let $k = 2$ and $p_1 = p_2 = \frac{1}{2}$. Suppose

$$S = \{\{x_1 + 2x_2 - 5\}, \{2x_1 + x_2 - 5\}\}.$$

If the outcomes of three die rolls are $1, 2, 1$, in that order, then $s_3 = (2, 1)$. Evaluating s_3 in S_2 yields $\{0\} \subset \mathbb{Z}_{\ge 0}$, while evaluating in S_1 yields $\{-1\}$. Nevertheless, $s_3 \in A_2$ and hence the game stops after 3 die rolls.

If instead $s_3 = (1, 2)$ then evaluating s_3 in S_1 yields $\{0\} \subset \mathbb{Z}_{\ge 0}$, and the goal is again reached.

3. Expected time to reach any goal

In this section, we describe a symbolic approach for studying the distribution of terminal states and the expected time to reach the goal defined in Definition 2.1.

For $s = (s_1, s_2, \ldots, s_k) \in \mathbb{Z}^k_{\ge 0}$ with $t = \sum_{i=1}^k s_i$, let

$$\mathbb{P}(s_t = s),$$

denote the multinomial probability of observing the sequence s after t die rolls with distribution P.

We emphasize that all generating functions in this paper encode the distribution of the terminal state s_τ, that is, the first state at which the process enters the target region A.

Definition 3.1 (Terminal state generating function)**.** Define the multivariate generating function

$$G(x_1, x_2 \ldots, x_k) \coloneqq \sum_{s \in A} \mathbb{P}\left(s_\tau = s\right) x_1^{s_1} x_2^{s_2} \cdots x_k^{s_k}.$$

The coefficient of $x_1^{s_1} x_2^{s_2} \cdots x_k^{s_k}$ in $G(x_1, x_2 \ldots, x_k)$ represents the probability that the game terminates for the first time at the state s.

Since the game terminates almost surely, we have

$$G(1, 1, \ldots, 1) = 1.$$

We begin with an example to illustrate the notation before introducing the general symbolic framework.

Example 2. Let $S = \{\}$ and $P = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. **Lucy: write example later**

For computational purposes, we impose a cutoff $K \in \mathbb{N}$ and define the truncated generating function,

$$G_K(x_1, x_2 \ldots, x_k) \coloneqq \sum_{\substack{s \in \mathbb{Z}_{\geq 0}^k \\ \sum_{i=1}^k s_i \leq K}} \mathbb{P}\left(s_\tau = s\right) x_1^{s_1} x_2^{s_2} \cdots x_k^{s_k}. \tag{1}$$

Note that G_K records the distribution of terminal states whose stopping time satisfies $\tau \leq K$. As $K \to \infty$, G_K converges coefficientwise to G.

All computations in this paper are carried out using $G_K(x_1, \ldots, x_k)$. These computations are implemented in a Maple package (see **??**). We now describe how to compute G_K explicitly. First we define an operator that extracts precisely those multinomial terms whose exponent sequences satisfy the goal condition in Definition 2.1.

Definition 3.2 (Operator)**.** For a polynomial

$$H(x_1, \ldots, x_k) = \sum_{e \in \mathbb{Z}_{\geq 0}^k} a_e x_1^{e_1} \cdots x_k^{e_k},$$

and a sequence $e = (e_1, e_2, \ldots, e_k)$, define

$$\mathcal{P}(H) \coloneqq \sum_{e \in A} a_e x_1^{e_1} \cdots x_k^{e_k},$$

where $A = \bigcup_{j=1}^{\ell} A_j$.

We provide two examples of how to use the operator.

Example 3. Lucy: finish this example later

Next, let $P(x_1, x_2, \ldots, x_k)$ be the probability generating function of the k-sided die:

$$P(x_1, x_2, \ldots, x_k) = \sum_{j=1}^{k} p_j x_j = p_1 x_1 + p_2 x_2 + \cdots + p_k x_k.$$

We define an auxiliary sequence of multinomials $S_K(x_1, x_2, \ldots, x_k)$ that takes care of the *survivors* at the Kth round. We initialize $S_0(x_1, x_2, \ldots, x_k) = 1$ and $G_0(x_1, x_2, \ldots, x_k) = 0$. Suppose that $G_{K-1}(x_1, x_2, \ldots, x_k)$ has been already computed. On the next die roll, with the previous survival multinomial, $S_{K-1}(x_1, x_2, \ldots, x_k)$, we define

$$N_K(x_1, x_2, \ldots, x_k) \coloneqq \mathcal{P}(P(x_1, x_2, \ldots, x_k) S_{K-1}(x_1, x_2, \ldots, x_k)), \tag{2}$$

$$S_K(x_1, x_2, \ldots, x_k) \coloneqq P(x_1, x_2, \ldots, x_k) S_{K-1}(x_1, x_2, \ldots, x_k) - N_K(x_1, x_2, \ldots, x_k), \tag{3}$$

$$G_K(x_1, x_2, \ldots, x_k) \coloneqq G_{K-1}(x_1, x_2, \ldots, x_k) + N_K(x_1, x_2, \ldots, x_k). \tag{4}$$

In Equation (2), $P(x_1, x_2, \ldots, x_k) S_{K-1}(x_1, x_2, \ldots, x_k)$ is the probability generating function, according to the sequences that record the count of appearances of each side of the die, where some of the multinomial terms might or might not satisfy the goal in Definition 2.1. By applying the operator $\mathcal{P}$, we extract the multinomial terms that satisfy the goal, all $s_t \in \cup_{j=1}^{\ell} A_j$. Hence, N_K are the multinomials that record the sequence that satisfies the goal on the K-th die roll. In Equation (3), S_K are the multinomials that record the sequences that do not yet satisfy the goal on the K-th die roll, and G_K as the total probability generating function of all terminal states reached in at most K rounds. We illustrate the algorithm for the first few steps in the following example.

Example 4. Let $k = 2$ and $p_1 = p_2 = \frac{1}{2}$. Suppose

$$S = \{\{2x_1 + x_2 - 6\}, \{x_1 + 2x_2 - 6\}\}.$$

Then, the probability generating function of the 2-sided die is

$$P(x_1, x_2) = \frac{1}{2}x_1 + \frac{1}{2}x_2.$$

On the first die roll,

$$N_1(x_1, x_2) = \mathcal{P}(P(x_1, x_2) \cdot S_0(x_1, x_2)) = \mathcal{P}((\tfrac{1}{2}x_1 + \tfrac{1}{2}x_2) \cdot 1) = 0.$$

Hence,

$$\begin{aligned} S_1(x_1, x_2) &= \tfrac{1}{2}x_1 + \tfrac{1}{2}x_2, \\ G_1(x_1, x_2) &= 0. \end{aligned}$$

Now, on the second die roll,

$$N_2(x_1, x_2) = \mathcal{P}(P(x_1, x_2) \cdot S_1(x_1, x_2)) = \mathcal{P}((\tfrac{1}{2}x_1 + \tfrac{1}{2}x_2) \cdot (\tfrac{1}{2}x_1 + \tfrac{1}{2}x_2)) = \mathcal{P}\left(\frac{1}{4}x_1^2 + \frac{1}{2}x_1x_2 + \frac{1}{4}x_2^2\right).$$

Notice that

$$\mathcal{P}\left(\frac{1}{4}x_1^2 + \frac{1}{2}x_1x_2 + \frac{1}{4}x_2^2\right) = 0,$$

since none of the sequences $(2, 0), (1, 1), (0, 2)$ have not entered the target regions.

Thus,

$$\begin{aligned} N_2(x_1, x_2) &= 0, \\ S_2(x_1, x_2) &= \frac{1}{4}x_1^2 + \frac{1}{2}x_1x_2 + \frac{1}{4}x_2^2 - 0 = \frac{1}{4}x_1^2 + \frac{1}{2}x_1x_2 + \frac{1}{4}x_2^2, \\ G_2(x_1, x_2) &= 0. \end{aligned}$$

On the third die roll,

$$\begin{aligned} N_3(x_1, x_2) = \mathcal{P}(P(x_1, x_2) \cdot S_2(x_1, x_2)) &= \mathcal{P}((\tfrac{1}{2}x_1 + \tfrac{1}{2}x_2) \cdot (\tfrac{1}{4}x_1^2 + \tfrac{1}{2}x_1x_2 + \tfrac{1}{4}x_2^2)) \\ &= \mathcal{P}\left(\frac{1}{8}x_1^3 + \frac{3}{8}x_1^2x_2 + \frac{3}{8}x_1x_2^2 + \frac{1}{8}x_2^3\right). \end{aligned}$$

Now,

$$\mathcal{P}\left(\frac{1}{8}x_1^3 + \frac{3}{8}x_1^2x_2 + \frac{3}{8}x_1x_2^2 + \frac{1}{8}x_2^3\right) = \frac{x_1^3}{8} + \frac{x_2^3}{8},$$

since $(3, 0)$ and $(0, 3)$ have entered at least one of the target regions, either $A_1 = \{x \in \mathbb{Z}_{\geq 0}^2 : f(x) \geq 0 \text{ for all } f \in S_1\}$ or $A_2 = \{x \in \mathbb{Z}_{\geq 0}^2 : f(x) \geq 0 \text{ for all } f \in S_2\}$. Indeed, evaluating $(3, 0)$ in $S_1 = \{2x_1 + x_2 - 6\}$ yields $\{0\} \subset \mathbb{Z}_{\geq 0}$, and evaluating $(0, 3)$ in $\{x_1 + 2x_2 - 6\}$ yields $\{0\} \subset \mathbb{Z}_{\geq 0}$. Therefore,

$$\begin{aligned} N_3(x_1, x_2) &= \frac{x_1^3}{8} + \frac{x_2^3}{8}, \\ S_3(x_1, x_2) &= \frac{1}{8}x_1^3 + \frac{3}{8}x_1^2x_2 + \frac{3}{8}x_1x_2^2 + \frac{1}{8}x_2^3 - N_3(x_1, x_3) = \frac{3}{8}x_1^2x_2 + \frac{3}{8}x_1x_2^2, \\ G_3(x_1, x_2) &= \frac{x_1^3}{8} + \frac{x_2^3}{8}. \end{aligned}$$

Continuing in this manner, with $K = 4$, we obtain

$$G_4(x_1, x_2) = \frac{1}{8}x_1^3 + \frac{1}{8}x_2^3 + \frac{3}{16}x_1^3 x_2 + \frac{3}{8}x_1^2 x_2^2 + \frac{3}{16}x_1 x_2^3.$$

This process continues until we reach the Kth round.

The probability that the game ends in at most K rounds is $G_K(1, 1, \ldots, 1)$. For sufficiently large K, this probability is arbitrarily close to 1. Also, the *conditional probability generating function* is

$$\bar{G}_K(x_1, x_2, \ldots, x_k) \coloneqq \frac{G_K(x_1, x_2, \ldots, x_k)}{G_K(1, 1, \ldots, 1)}.$$

4. Recurrences for special cases

In this section, we present linear recurrences for different cases of reaching a predetermined goal.

Definition 4.1. Let p_i be the probability of the die landing on side i. Denote $a_k(n;\ p_1, p_2, \ldots, p_k)$ to be the expected number of rolls of a k-sided die until n occurrences of one of the sides are seen. Furthermore, let

$$a(n) \coloneqq a_k(n;\ p_1, p_2, \ldots, p_k),$$

whenever k and $p_1, p_2, \ldots, p_k$ are clear from context. If $k = 2$, then the notion of throwing a die reduces to tossing a coin.

We begin by providing a recurrence for $k = 2$ such that with probability p the coin lands on heads and with probability $q = 1 - p$ the coin lands on tails. Then,

$$a(n+2) = \frac{-4np^2 + 4np - 2p^2 + n + 2p + 2}{n+1} a(n+1) - \frac{2q(2n+1)p}{n} a(n), \tag{5}$$

with $a(1) = 1, a(2) = -2p^2 + 2p + 2$ and $a(n) = a_2(n;\ p, q)$.

While finding a general solution to Equation 5 might not be feasible by hand analysis, the recurrence enables fast computations for any p and n.

Next, we provide a recurrence for $k = 3$ where each side of the die has equal probability, $p_1 = p_2 = p_3 = \frac{1}{3}$, then

$$\begin{aligned} a(n+3) = {} & \frac{486n^3 + 2079n^2 + 2853n + 1198}{18\,(9n+7)\,(2+n)^2} a(n+2) \\ & - \frac{\left(243n^4 + 1161n^3 + 2043n^2 + 1585n + 458\right)}{9\,(n+1)\,(9n+7)\,(n+2)^2} a(n+1) \\ & + \frac{(3n+2)\,(2n+1)\,(3n+1)\,(9n+16)}{18\,(9n+7)\,(n+2)^2\, n} a(n) \end{aligned} \tag{6}$$

with $a(1) = 1, a(2) = \frac{26}{9}, a(3) = \frac{409}{81}$, and $a(n) = a_3(n;\ \frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.

For the case where $k = 4$ where each side has equal probability, $p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$, we get a fifth-order recurrence. We first define the following expressions,

$$\begin{aligned}
Y &= 28152901853008477n^2 + 81763683068059260n + 53615706600102264, \\
A &= 5514198858742237n^2 + 22377255642631594n + 21709635674546661, \\
B &= 14460392818937498400n^7 + 182606360698176085584n^6 + 1052157967879148282152n^5 \\
&\quad + 3484773784146230875959n^4 + 6979199009322021232333n^3 + 8314830354132186721782n^2 \\
&\quad + 5412609663886380033970n + 1480437544492104202620, \\
C &= 839213571251447020800n^8 + 144363492982916631475 2n^7 + 11054199574946609460984n^6 \\
&\quad + 48676841768444583401754n^5 + 133788726394565359658831n^4 + 233751077270618823645175n^3 \\
&\quad + 252381873231581655737525n^2 + 153212526518476108150989n + 39775428194986725479910, \\
D &= 1100011429745394105600n^7 + 19152792386264112468480n^6 + 143043929859374673032000n^5 \\
&\quad + 589761825668127573937780n^4 + 1440262566068342887082250n^3 + 2068390811776651286642870n^2 \\
&\quad + 1602213872244207150492700n + 509319076750761767593320 \\
E &= 34020232205983529760n^5 + 45087003833845691001 6n^4 + 232747114169276702697 4n^3 \\
&\quad + 5760259258259629592377n^2 + 6685431166857223519260n + 2835874645996633400088.
\end{aligned}$$

Then,

$$\begin{aligned}
a(n+5) = {} & \frac{E}{Y}\frac{1}{288(n+4)^3}a(n+4) - \frac{D}{Y}\frac{1}{576(n+3)^2(n+4)^3}a(n+3) \\
& + \frac{C}{Y}\frac{1}{576(n+2)(n+3)^2(n+4)^3}a(n+2) - \frac{B}{Y}\frac{(2n+3)}{576(n+2)(n+3)^2(n+4)^3}a(n+1) \\
& + \frac{A}{Y}\frac{(4n+1)(3n+2)(2n+3)(2n+1)(3n+1)(4n+3)}{576(n+2)(n+3)^2(n+4)^3}a(n), \qquad (7)
\end{aligned}$$

where $a(1) = 1, a(2) = \frac{103}{32}, a(3) = \frac{48039}{8192}, a(4) = \frac{2288659}{262144}, a(5) = \frac{6301126135}{536870912}$, and $a(n) = a_4(n;\ \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$.

Acknowledgements

L. Martinez was supported by the NSF Graduate Research Fellowship Program under Grant No. 2233066 and in part by a Joel Lebowitz Summer Research Fellowship.

References

[1] Noga Alon and Yaakov Malinovsky, *Hitting a prime in 2.43 dice rolls (on average)*, Amer. Statist. **77** (2023), no. 3, 301–303.

[2] Noga Alon, Yaakov Malinovsky, Lucy Martinez, and Doron Zeilberger, *Hitting k primes by dice rolls*, Electron. J. Combin. **32** (2025), no. 4, Art. P4.16.

[3] Simon Blatt, Uta Freiberg, and Vladimir Shikhman, *Mathematics of family planning in talmud*, Math. Intell. **47** (2025), 120–125.

[4] Shane Chern, *Hitting a prime by rolling a die with infinitely many faces*, Amer. Statist. **78** (2024), no. 3, 297–303.

[5] A. W. F. Edwards, *Pascal and the problem of points*, Int. Stat. Rev. **50** (1982), no. 3, 259–266.

[6] William Feller, *An Introduction to Probability Theory and its Applications*, Vol. 1, John Wiley & Sons, New York-London-Sydney, 1991.

[7] Svante Janson, Lucy Martinez, and Doron Zeilberger, *How many coin tosses would you need until you get n Heads or m Tails?*, arXiv preprint (2025), available at `arXiv:2512.07803`.

[8] Lucy Martinez and Doron Zeilberger, *How many dice rolls would it take to reach your favorite kind of number?*, Maple Transactions **3** (2023), no. 3, Art. 15953.

[9] Jonathan Rosenberg, *Some examples of mathematical analysis applied to talmud study.* Available at https://math.umd.edu/ jmr/MathTalmud.html, based on a talk given at KMS, January 20, 2001.

(L. Martinez) Department of Mathematics, Rutgers University, Piscataway, NJ 08854
Email address: `lucy.martinez@rutgers.edu`