Problems:

1. How many die do you have to roll before you get a `5` followed by a `5`?
2. How about for a `5` followed by a `6`?

Solution:

We're going to solve the problem by calculating a weighted average:

$$\langle \ell \rangle = \frac{\sum_\ell \ell \times p(\ell)}{\sum_\ell p(\ell)}. \tag{1}$$

The trouble is we don't have the length probability distribution, $p(\ell)$, and we need to restrict the sum to roll sequences that end in `55` (or `56`). But we don't need to write out these sequences by hand.

To get a feel for why, let's try a simpler version of the `55` problem: how many times do we need to flip a coin before we get `T` followed by `T`?

1 The `TT` problem

It's easy enough to write down the first few possibilities: `TT`, `HTT`, `HHTT`, `THTT`, `HHHTT`, `THHTT`, `TTHTT`, `HTHTT`, `THTHTT`, ...

If we had the full sequence, the average number of flips would be

$$\langle \ell_{\text{ends in }\mathtt{TT}} \rangle = \frac{2 \times P(\mathtt{TT}) + 3 \times P(\mathtt{HTT}) + 4 \times (P(\mathtt{HHTT}) + P(\mathtt{THTT})) + \dots}{P(\mathtt{TT}) + P(\mathtt{HTT}) + P(\mathtt{HHTT}) + P(\mathtt{THTT}) + \dots} \tag{2}$$

The bottom is just the sum of the probabilities of all possible outcomes and the top is the sum over all outcomes of the number of flips times its probability of occurrence.

But it's nerve-wracking to enumerate these sequences by hand — is there a case we forgot, did we double count something, is there an internal `TT` we missed somewhere?

We can alleviate these concerns by generating them automatically, we just have to come up with an appropriate symbol alphabet.

All the permissible flip sequences are `TT`-free "**stems**" capped by `TT`. A typewriter with keys for `H` and `TH` could therefore generate all the stems for these sequences. We'd just have to cap them by tacking a `TT` on at the end.

> **Exercise:** Convince yourself that an {`H, TH`} keyboard can generate all permissible sequences, or try to find a counterexample.

Representing the typewriter keys by their probabilities, the stems that can be generated by one keypress are $g = (P_{\mathtt{H}} + P_{\mathtt{TH}})$ which become $P_{\mathtt{H}} \cdot P_{\mathtt{TT}} + P_{\mathtt{TH}} \cdot P_{\mathtt{TT}}$ when we cap them. Roll sequence probabilities concatenate like $P_{\mathtt{X}} \cdot P_{\mathtt{Y}} = P_{\mathtt{XY}}$, so the single keypress roll sequences become $P_{\mathtt{HTT}} + P_{\mathtt{THTT}}$.

The stems generated by two keypresses are

$$g^2 = (P_{\mathtt{H}} + P_{\mathtt{TH}})^2 = P_{\mathtt{HH}} + P_{\mathtt{HTH}} + P_{\mathtt{THH}} + P_{\mathtt{THTH}}$$

which become $P_{\mathtt{HHTT}} + P_{\mathtt{HTHTT}} + P_{\mathtt{THHTT}} + P_{\mathtt{THTHTT}}$ when we cap them[1].

> **Are you awake:** Find the sum of probabilities for all possible stems (i.e. the denominator to Eq. 5) in terms of g. *Hint:* don't forget the sequence `TT` which corresponds to pressing no keys on the stem generating typewriter.

[1]The probability of any particular sequence is just the product of the letter probabilities $P_{\mathtt{THTHTT}} = P_{\mathtt{T}} \cdot P_{\mathtt{H}} \cdot P_{\mathtt{T}} \cdot P_{\mathtt{H}} \cdot P_{\mathtt{T}} \cdot P_{\mathtt{T}}$. For coin flips, this is just $1/2^{\ell}$ but we'll keep it written in per-symbol probabilities looking forward to dice.

The sum of all possible sequence probabilites is generated by

$$G = \sum_{n=0}^{\infty} \left(P_{\mathtt{H}} + P_{\mathtt{TH}}\right)^n P_{\mathtt{TT}} = \Big[1 + \left(P_{\mathtt{H}} + P_{\mathtt{TH}}\right) + \left(P_{\mathtt{H}} + P_{\mathtt{TH}}\right)^2 + \cdots\Big] \cdot P_{\mathtt{TT}},$$

which gets us the sum of probabilites in the denominator of Eq. 1, but how can we generate the weighted sum in the numerator?

We need to keep track of how many symbols are in the sequence corresponding to each product of keypresses (e.g. $P_{\mathtt{H}} \cdot P_{\mathtt{TH}} \cdot P_{\mathtt{TT}}$). If we attach one factor of a counter z to each $P_{\mathtt{H}}$ and two factors of z to each $P_{\mathtt{TH}}$ and $P_{\mathtt{TT}}$ in G, then the power of z on the terms of the expanded series will be equal to the length of the corresponding sequence.

In other words,

$$\begin{aligned} G(z) &= \Big[1 + \left(P_{\mathtt{H}} \cdot z + P_{\mathtt{TH}} \cdot z^2\right) + \left(P_{\mathtt{H}} \cdot z + P_{\mathtt{TH}} \cdot z^2\right)^2 + \cdots\Big] \cdot P_{\mathtt{TT}} \cdot z^2 \\ &= P_{\mathtt{TT}} z^2 + P_{\mathtt{HTT}} z^3 + \left(P_{\mathtt{HHTT}} + P_{\mathtt{THTT}}\right) z^4 + \left(P_{\mathtt{HHHTT}} + \cdots\right) z^5 \ldots \end{aligned}$$

Exercise: Calculate and interpret $G'(z)$.

So, $G(z)$ can get us both the numerator and denominator of the weighted sum. $G(z)$ evaluated at $z = 1$ gets us the sum of probabilities, and its derivative gets us the weighted sum of the number of flips we wanted in Eq. 1.

We can tidily write the average number of rolls like

$$\begin{aligned}\langle \ell_{\text{ends in }\texttt{55}} \rangle &= \partial_z \log G(z)|_{z=1} \\ &= \left.\frac{G'(z)}{G(z)}\right|_{z=1}.\end{aligned}$$

All we need to do now is to find a closed expression for $G(z)$.

From above, $G(z)$ is just a geometric series in $\left(P_{\texttt{H}} \cdot z + P_{\texttt{TH}} \cdot z^2\right)$, so:

$$G(z) = \sum_{n=0}^{\infty} \left(P_{\texttt{H}} \cdot z + P_{\texttt{TH}} \cdot z^2\right)^n P_{\texttt{TT}} \cdot z^2 \tag{3}$$

$$= \frac{P_{\texttt{TT}} z^2}{1 - \left(P_{\texttt{H}} \cdot z + P_{\texttt{TH}} \cdot z^2\right)}. \tag{4}$$

So the expected number of flips to get TT is just

$$\begin{aligned}\langle \ell_{\text{ends in }\texttt{TT}} \rangle &= \partial_z \log G(z)|_{z=1} \\ &= \left.\frac{1}{G(z)} \partial_z G(z)\right|_{z=1} \\ &= \frac{1}{G(z)} \times \left.\frac{2P_{\texttt{TT}} \cdot z \cdot \left(1 - P_{\texttt{H}} \cdot z - P_{\texttt{TH}} \cdot z^2\right) + \left(P_{\texttt{H}} + 2P_{\texttt{TH}} \cdot z\right) P_{\texttt{TT}} \cdot z^2}{\left[1 - \left(P_{\texttt{H}} \cdot z + P_{\texttt{TH}} \cdot z^2\right)\right]^2}\right|_{z=1} \\ &= \frac{2P_{\texttt{TT}} - 2P_{\texttt{TT}} P_{\texttt{H}} - 2P_{\texttt{TT}} P_{\texttt{TH}} + P_{\texttt{H}} P_{\texttt{TT}} + 2P_{\texttt{TH}} P_{\texttt{TT}}}{P_{\texttt{TT}} \left[1 - \left(P_{\texttt{H}} + P_{\texttt{TH}}\right)\right]} \\ &= \frac{1 + \left(1 - P_{\texttt{H}}\right)}{1 - \left(P_{\texttt{H}} + P_{\texttt{TH}}\right)} \qquad (5)\end{aligned}$$

Plugging in $P_{\texttt{H}} = 1/2$ and $P_{\texttt{TH}} = 1/4$, we get

$$\langle \ell_{\text{ends in }\texttt{TT}} \rangle = 6.$$

> **Exercise:** What would happen if we left off the cap term, $P_{\texttt{TT}} \cdot z^2$, in the generating function $G(z)$?

Exercise: Interpret the expression for $\langle \ell_{\text{ends in } \texttt{TT}} \rangle$.

Just to make sure all went well, we can plot the coefficients of $G(z)$ against a 250,000 run simulation:

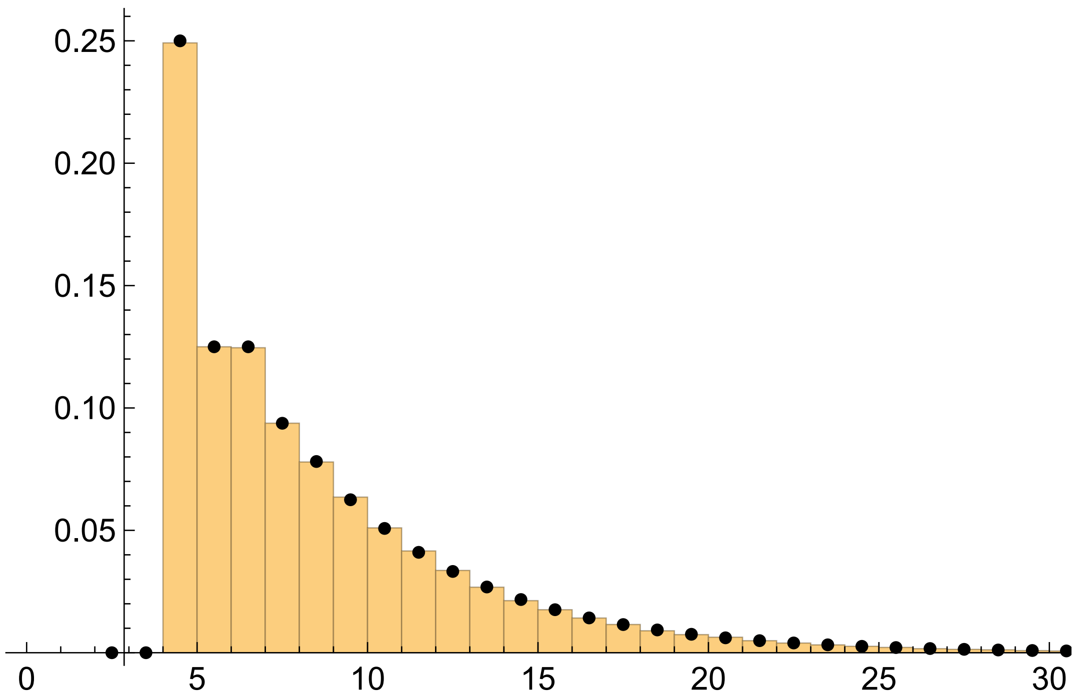

Figure 1: $[z^n]G(z)$ vs empirical probabilities from simulation

2 The 55 problem

With that, we have all the tools we need to solve the 55 dice problem. What's that? Whoops too bad, we already solved it. Let's see how.

The valid sequences for the 55 problem are all sequences of rolls that contain no internal 55 and are capped by 55. Its stem-generating typewriter therefore has keys for `1, 2, 3, 4, 6, 51, 52, 53, 54,` and 56.[2]

The corresponding generating term is

$$g(z) = \left[(P_1 + P_2 + P_3 + P_4 + P_6) \cdot z + (P_{51} + P_{52} + P_{53} + P_{54} + P_{56}) \cdot z^2\right]$$

But we can just write the first sum of probabilities as $P_{\sim 5}$ and the second

[2]This typewriter can generate all possible sequences that have no internal repeated 5s.

as $P_{\texttt{5X|X} \neq \texttt{5}}$ so the generating term looks like

$$g(z) = \left(P_{\sim\texttt{5}} \cdot z + P_{\texttt{5X|X} \neq \texttt{5}} \cdot z^2\right)$$

which looks the same as the generating term in Eq 3.

So we can just replace $P_{\texttt{H}} \to P_{\sim\texttt{5}}$, $P_{\texttt{TH}} \to P_{\texttt{5X|X}\neq\texttt{5}}$, and the cap $P_{\texttt{TT}} \to P_{\texttt{55}}$ in Eq. 5:

$$\langle \ell_{\text{ends in }\texttt{55}} \rangle = \frac{1 + (1 - P_{\sim\texttt{5}})}{1 - (P_{\sim\texttt{5}} + P_{\texttt{5X|X}\neq\texttt{5}})}$$

Plugging in $P_{\sim\texttt{5}} = 5/6 = 30/36$ and $P_{\texttt{5X|X}\neq\texttt{5}} = 5/36$, we get

$$\langle \ell_{\text{ends in }\texttt{55}} \rangle = 42$$

3 The `56` problem

`56` capped sequences are a little bit harder, which we can see by trying to build a stem-generating typewriter for them. For the `55` keyboard, we withheld the `5` key and replaced it with five custom bi-symbol keys to ensure that all internal `5`s would be followed by a non-`5`.

The naive extension of that keyboard to the `56` case would be to remove the `56` key and add one for `55` so that the total keyset becomes:

$$\{\texttt{1, 2, 3, 4, 6, 51, 52, 53, 54, 55}\}.$$

Exercise: Does this typewriter generate **any** impermissible stems?

Exercise: Does this typewriter generate **all** permissible stems?

This keyboard has two dueling constraints:

1. strings of repeated `5`s are permissible if they don't end in `6`, and
2. a `6` must be able to appear anywhere except directly after a `5`.

To accomodate strings of repeated `5`s while protecting them from contiguous `6`s, we have to generalize the bi-symbol keys from the `55` keyboard to tri-symbols, quad-symbols, quint-symbols, and even sex-symbols.

Actually we have to go all the way to ∞-symbols. The rows of our keyboard now include

- `1, 2, 3, 4, 6`
- `51, 52, 53, 54`
- `551, 552, 553, 554`
- `5551, 5552, 5553, 5554`
- ...

The multi-symbol rolls seem like they defeat the purpose of the generating function, which is to automatically generate the infinite set of outcomes by writing down a finite set of keys. But they are no bother because they can themselves be tidily generated:

$$\begin{aligned} P_{\mathtt{51}} \cdot z^2 + P_{\mathtt{551}} \cdot z^3 + P_{\mathtt{5551}} \cdot z^4 + \ldots &= P_{\mathtt{5}} P_{\mathtt{1}} \cdot z^2 + P_{\mathtt{5}} P_{\mathtt{5}} P_{\mathtt{1}} \cdot z^3 + P_{\mathtt{5}} P_{\mathtt{5}} P_{\mathtt{5}} P_{\mathtt{1}} \cdot z^4 + \ldots \\ &= \left(P_{\mathtt{5}} \cdot z + P_{\mathtt{5}} P_{\mathtt{5}} \cdot z^2 + P_{\mathtt{5}} P_{\mathtt{5}} P_{\mathtt{5}} \cdot z^3 + \ldots \right) P_{\mathtt{1}} \cdot z \\ &= \left(\frac{1}{1 - P_{\mathtt{5}} \cdot z} - 1 \right) P_{\mathtt{1}} \cdot z \end{aligned}$$

> **Are you awake:** Why is there a -1 in the last line?

Adding in the sequences that end in in `2`, `3`, and `4`, we get

$$\left(\frac{1}{1 - P_{\mathtt{5}} \cdot z} - 1 \right) \left(P_{\mathtt{1}} \cdot z + P_{\mathtt{2}} \cdot z + P_{\mathtt{3}} \cdot z + P_{\mathtt{4}} \cdot z \right) = \left(\frac{1}{1 - P_{\mathtt{5}} \cdot z} - 1 \right) P_{\mathtt{X}} \cdot z$$

where $P_{\mathtt{X}} = P_{\mathtt{1}} + P_{\mathtt{2}} + P_{\mathtt{3}} + P_{\mathtt{4}}$.

With that, we nearly have a complete set of keys for our typewriter, and the stem generating term is

$$g(z) = [\overbrace{P_{\sim 5} \cdot z}^{\text{1,2,3,4,6}} + \overbrace{\left(\frac{1}{1 - P_{\mathtt{5}} \cdot z} - 1 \right) P_{\mathtt{X}} \cdot z]}^{\text{51,52,53,54,551,552,553,554,...}}$$

To get the generating function for all possible sequences, we just add up the terms for zero, one, two, three, etc. keypresses, and cap them all off with `56`:

$$\begin{aligned} G(z) &= \left[1 + g(z) + g(z)^2 + g(z)^3 + \ldots\right] P_{\texttt{56}} \cdot z^2 \\ &= \left[\sum_{n=0}^{\infty} g(z)^n\right] P_{\texttt{56}} \cdot z^2 \\ &= \frac{1}{1 - g(z)} P_{\texttt{56}} \cdot z^2 \end{aligned}$$

> **Exercise:** Find a valid roll sequence that this typewriter **can't** generate.

The keyboard we've built can generate almost every sequence that's allowed in the `56` case. The only sequences it can't generate are ones that end with a string of `5s` followed immediately by a `6`.

If we included keys for uncapped strings of `5`s in the stem-generating keyboard, they would work with the `6` key to generate sequences with impermissible internal `56`s.

But if we make a second keyboard with keys for {`-, 5, 55, 555, ...`}, that we can press at most once after we finish with the first keyboard, then we'll generate all permissible roll sequences.

The complete generating function $G(z)$ is

$$G(z) = \overbrace{\frac{1}{1 - g(z)}}^{\text{Keyboard 1}} \overbrace{\frac{1}{1 - P_{\texttt{5}} \cdot z}}^{\text{Keyboard 2}} \overbrace{P_{\texttt{56}} \cdot z^2}^{\text{Cap}}.$$

> **Exercise:** Calculate $G'(z)|_{z=1}$ and show that $\langle \ell_{\text{ends in } \texttt{56}} \rangle = 36$.

> **Exercise:** Interpret the result for $\langle \ell_{\text{ends in } \texttt{56}} \rangle$.

4 Why not just use expectations?

The problems we did here could all have been done by finding relationships between expectation values — for the `55` problem, the expected number of rolls can be found by just thinking about the three things that could happen.

We could:

1. immediately roll a `56`, or
2. roll not a 5, and start at square one having wasted a roll, or
3. roll a `5`, then not a `5`, then start at square one having wasted two rolls

this yields

$$\langle \ell \rangle = 2 \times \frac{1}{36} + \frac{30}{36} \left(\langle \ell \rangle + 1 \right) + \frac{5}{36} \left(\langle \ell \rangle + 2 \right)$$

which gets $\langle \ell \rangle = 42$.

This is fine if we want to calculate averages but it falls flat for finding other statistics, which is sad. Probabilistic systems are inherently stochastic and we want to know whether the average we calculate is the center of a tight peak, or the diffuse consensus of a wide distribution.

Might the generating function tree bear fruit a second time? Let's see what wonders it has in store.

Important Exercise: Calculate $\partial_z^2 \log G(z)|_{z=1}$ and interpret its significance.

This result, while startling, is a little underwhelming. Couldn't we have a clean connection between higher moments and derivating our generating function instead of adding $\partial_z \log G(z) + \partial_z^2 \log G(z)$?

In fact, this is but a mirage brought unto us by our own lack of foresight. Each time we take a derivative of z^n, we alter its nature, and the next time it is but a husk of its original self. Had we been wise enough to employ e^z as our counter variable instead of z, might things have turned out differently?

Exercise: Replace z with e^z in the generating function and take the derivative. What must we set z to this time?

If that seems to be working, it's time to take another crack at finding higher moments.

Important Exercise: Calculate $\partial_z^2 \log G(e^z)$ and evaluate it at the value of z you found above.

Exercise: Use your last result to show that $\sigma^2(\ell_{\text{ends in TT}})$ is 22.

The derivative properties of $\log G(e^z)$ are a wonder to behold and are testament to the wider realm of magical things you can do with generating functions. Here's a nice problem to leave off:

Exercise: Suppose you're polling a population on a topic where a fraction p_A of the citizens believes one thing and $p_B = 1 - p_A$ believe the other. Find the generating function for the possible results of polling n citizens at random.

Exercise: Taking the expected polling error to be the variance, find the expected margin of error for the poll.