Double Heads

Flip a coin until two consecutive heads appear. Assume that the coin flips are all independent. and that the coin comes up heads with probability $0 < p < 1$.

Let H_i be the indicator of heads on the ith roll, for $i \in \mathbb{Z}_+$.

Let N be the number of flips (inclusive) until heads first appears on two consecutive flips.

Find the PMF of N, p_N.

This is an interesting problem in many ways. The naive approach would be make a direct analogy to the modem connection problem. We might reason as follows: "Consider pairs of flips. Heads followed by heads occurs with probability p^2 and not with probability $1-p^2$. Using the modem connection problem solution, we have that the probability that we get two consecutive heads on the first time in $2n$ flips is $(1-p^2)^{n-1}p^2$." Unfortunately, this reasoning is wrong. Why?

A more careful analysis of the sequences that can occur tells us that the observed sequences consist of a series of `T`s and `HT`s followed by `HH`. (In regular expression terms, this is `(T|HT)*HH`.) We can use the modem connection problem here in one way. Since `T` occurs with probability $1-p$, `HT` with probability $p(1-p)$, and `HH` with probability p^2 (the three terms sum to 1), the probability that we see `HH` after k `T`s and `HT`s is $(1-p^2)^k p^2$. This is right, but unfortunately, it answers the wrong question. To determine the number of *flips*, we have to carefully count all the ways that `T`s and `HT`s can occur in this sequence. Messy!

Generating functions give us a fast and generalizable approach. Consider all the sequences that can possibly be observed in this experiment. We can express this as a formal sum that simply accumulates the possibilities

$$\mathsf{G}_N = \mathtt{HH} + \mathtt{THH} + \mathtt{HTHH} + \mathtt{TTHH} + \mathtt{HTTHH} + \cdots.$$

Now, replace every `HT` by $p(1-p)z^2$, every `HH` by p^2z^2 and every remaining `T` by $(1-p)z$. Then, when we collect terms with common powers of z, we find that the coefficient of z^k is just $\mathsf{p}_N(k)$.

Define the *probability generating function* G_N by

$$\mathsf{G}_N(z) = \sum_{k\in\mathbb{Z}} \mathsf{p}_N(k)\, z^k.$$

This is a packaging of the PMF of N as a formal power series that we can manipulate somewhat differently.

Notice that

$$\mathsf{G}_N(1) = \sum_{k=2}^{\infty} \mathsf{p}_N(k) = 1,$$

and

$$\mathsf{G}'_N(1) = \sum_{k=2}^{\infty} k\mathsf{p}_N(k) = \mathsf{E}N.$$

So, if we can compute G_N, we can easily find many quantities of interest. (In fact, we can recover p_N. Do you see how?)

Going back to the formal sum above we have that every final sequence starts with T every

$$\begin{aligned}\mathsf{G}_N &= \mathtt{HH} + \mathtt{THH} + \mathtt{HTHH} + \mathtt{TTHH} + \mathtt{HTTHH} + \cdots \\ &= \mathtt{HH} + \mathtt{T}(\mathtt{HH} + \mathtt{THH} + \cdots) + \mathtt{HT}(\mathtt{HH} + \mathtt{THH} + \cdots) \\ &= \mathtt{HH} + \mathtt{T}\mathsf{G}_N + \mathtt{HT}\mathsf{G}_N.\end{aligned}$$

Or to put this more formally, for every z we have

$$\mathsf{G}_N(z) = p^2 z^2 + (1-p)z\mathsf{G}_N(z) + p(1-p)z^2\mathsf{G}_N(z).$$

Solving for $\mathsf{G}_N(z)$, we have

$$\mathsf{G}_N(z) = \frac{p^2 z^2}{1 - qz - pqz^2},$$

where $q = 1 - p$.

By taking $\mathsf{G}'_N(1)$, we can easily find that that $\mathsf{E}N = (1+p)/p^2$. Nifty.

In the case $p = 1/2$, we can write $\mathsf{G}_N(z) = (z/2)F(z/2)$, where

$$F(z) = \frac{z}{1 - z - z^2},$$

the generating function for the Fibonacci numbers $(0, 1, 1, 2, 3, 5, 8, 13, \ldots)$

$$F_k = 1_{(k=1)} + F_{k-1} + F_{k-2}.$$

It follows that

$$\mathsf{G}_N(z) = \sum_{k=2}^{\infty} \frac{F_{k-1}}{2^k} z^k,$$

so

$$\mathsf{p}_N(k) = \begin{cases} \frac{F_{k-1}}{2^k} & \text{if } k \in \mathbb{Z}_+ \\ 0 & \text{otherwise.} \end{cases}$$