1 Waiting for patterns

- Expected waiting time for patterns in Bernoulli trials

Suppose $X_1, X_2, \ldots$ are independent coin tosses with $\mathbb{P}(X_i = H) = p$, $\mathbb{P}(X_i = T) = 1 - p = q$. Take a particular pattern of some finite length $K$, say $\underbrace{HH \ldots H}_K$ or $\underbrace{HH \ldots H}_{K-1}T$. Let

$$T_{\text{pat}} := \text{first } n \text{ s.t. } (X_{n-K+1}, X_{n-K+2}, \ldots, X_n) = \text{pattern}$$

You can try to find the distribution of $T_{\text{pat}}$, but you will find it very difficult. We can compute $\mathbb{E}[T_{\text{pat}}]$ with our tools. Start with the case of $\text{pat} = \underbrace{HH \ldots H}_K$.

- For $K = 1$,

$$\mathbb{P}(T_H = n) = q^{n-1}p, \quad n = 1, 2, 3 \ldots$$

$$\mathbb{E}(T_H) = \sum_n nq^{n-1}p = \frac{1}{p}$$

For a general $K$, notice that when the first $T$ comes (e.g. $HHHT$, considering some $K \geq 4$), we start the counting process again with no advantage.

Let $m_K$ be the expected number of steps to get $K$ heads in a row. Let $T_T$ be the time of the first tail.

Observe:
If $T_T > K$, then $T_{HH\ldots H} = K$;
if $T_T = j \leq K$, then $T_{HH\ldots H} = j + \text{ (fresh copy of) } T_{HH\ldots H}$. 

Therefore $\mathbb{E}(T_{HH...H}|T_T > K) = K$ and $\mathbb{E}(T_{HH...H}|T_T = j) = j + m_K$. So condition $m_K$ on $T_T$:

$$m_K = \sum_{j=1}^{K} \mathbb{P}(T_T = j)(j + m_K) + \mathbb{P}(T_T > K)K$$

$$= \sum_{j=1}^{K} p^{j-1}q(j + m_k) + p^K K$$

$$= (\sum_{j=1}^{K} p^{j-1}q)m_K + (\sum_{j=1}^{K} p^{j-1}qj) + p^K K$$

$$= (1 - p^K)m_K + \mathbb{E}(T_T 1(T_T \leq K)) + p^K K$$

Recall that $\mathbb{E}(X 1_A) = \mathbb{E}(X|A)\mathbb{P}(A)$, and notice that $\mathbb{E}(T_T) = 1/q$, $\mathbb{E}(T_T|T_T > K) = K + 1/q$, so

$$\mathbb{E}(T_T 1(T_T \leq K)) = \mathbb{E}(T_T) - \mathbb{E}(T_T 1(T_T > K))$$

$$= \frac{1}{q} - (K + \frac{1}{q})p^K$$

Therefore

$$m_K = (1 - p^K)m_K + \frac{1}{q} - (K + \frac{1}{q})p^K + p^K K$$

$$= \frac{1 - p^K}{1 - p} \frac{1}{p^K}$$

$$= 1 + p + p^2 + \cdots + p^{K-1}$$

or

$$m_K p^K = 1 + p + p^2 + \cdots + p^{K-1}.$$

- Try to understand equation $m_K p^K = 1 + p + p^2 + \cdots + p^{K-1}$.

**Idea:** Imagine you are gambling and observe the sequence of $H$ and $T$ evolving. Each time you bet that the next $K$ steps will be $\overbrace{HH\ldots H}^{K}$. Scale to get $\$1$ if you win.

**Accounting:** Suppose that $K = 3$. 

Expected cost = $p^K m_K$
Expected return = $1 \cdot (1 + p + \ldots + p^{K-1})$
So by “Conservation of Fairness”, $m_K p^K = 1 + p + p^2 + \ldots + p^{K-1}$. Rigorously, this is a martingale argument, but details are beyond scope of this course. See "A Martingale Approach to the Study of Occurrence of Sequence Patterns in Repeated Experiments" by Shuo-Yen Robert Li, Ann. Probab. Volume 8, Number 6 (1980), 1171-1176.

- Try pattern = \textit{HTHT}

Again, make fairness argument:

\[ p^2 q^2 m_{\text{HTHT}} = 1 + pq \Rightarrow m_{\text{HTHT}} = \frac{1 + pq}{p^2 q^2} \]

- Try

Then \( m_{\text{HTTHHTT}} = \frac{1 + p^2 q^2}{p^4 q^4} \).
To verify the method described above actually works for all patterns of length $K = 2$, make a M.C. with states $\{HH, TT, TH, TT\}$ by tracking the last two states.

<table>
<thead>
<tr>
<th></th>
<th>HH</th>
<th>HT</th>
<th>TH</th>
<th>TT</th>
</tr>
</thead>
<tbody>
<tr>
<td>HH</td>
<td>p</td>
<td>q</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HT</td>
<td>0</td>
<td>0</td>
<td>p</td>
<td>q</td>
</tr>
<tr>
<td>TH</td>
<td>p</td>
<td>q</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TT</td>
<td>0</td>
<td>0</td>
<td>p</td>
<td>q</td>
</tr>
</tbody>
</table>

We have a chain with states $i \in \{HH, HT, TH, TT\}$, and we want the mean first passage time to HT. Say we start at $i = TT$, and define

$$m_{ij} = \text{mean first passage time to } (j = HT)$$

In general, for a M.C. with states $i, j, \ldots$,

$$m_{ij} = \mathbb{E}_i(\text{ time to reach } j)$$

If $i \neq j$,

$$m_{ij} = \sum_{k \neq j} P(i, k)(1 + m_{kj}) + \sum_{k = j} 1 \cdot P(i, j)$$

$$= 1 + \sum_{k \neq j} P(i, k)m_{kj}$$

**Remark:** The derivation shows that $j$ need not be an absorbing state.

So we want

$$m_{TT,HT} = 1 + p \ m_{TH,HT} + q \ m_{TT,HT}$$
$$m_{TH,HT} = 1 + p \ m_{HH,HT}$$
$$m_{HH,HT} = 1 + p \ m_{HH,HT}$$

Solve the system of equations and verify that $m_{TT,HT} = \frac{1}{pq}$. In previous notation, this is just $m_{HT}$, and the general formula is seen to be working. You can check similarly that the formula works for patterns of length 3 by considering a chain with 8 patterns as its states, and so on (in principle) for patterns of length $K$ with $2^K$ states.