ex. (length of the longest run of heads in successive coin flips)

A (poss. biased) coin is flipped n times. If

$$P(H) = p$$

what is

$$P(\text{there is a string of } k\ H\text{'s}) \ ?$$

A: Let H_i = event that the k flips i through $i+k-1$ are all H, so that what we are looking for is

$$P(\underbrace{H_i \text{ for at least one } i = 1, \dots, n-k+1}_{\text{call this } E})$$

Of course,

$$P(H_i) = p^k$$

but that is no comfort by itself, for there are $n-k+1$ starting positions of length-k strings & many of them overlap.

Let
$$\boxed{F_j = \text{event that the first tail appears at flip } j, \quad j = 1, \dots, k}$$

and observe that

$$H_1 \cap F_j = \emptyset \quad \text{for all } j = 1, \dots, k$$

and

$$F_j \cap F_k = \emptyset \quad \text{for } j \neq k$$

so that

$$\boxed{S = H_1 \sqcup F_1 \sqcup \cdots \sqcup F_k}$$

By the multiplication principle, therefore,

$$\boxed{P(E) = \underbrace{P(E|H_1)}_{1} P(H_1) + \sum_{j=1}^{k} P(E|F_j)P(F_j)}$$

* has prob. 1

Let us introduce some simplifying notation: for any $m > k$, write

$$\boxed{P_m \stackrel{\text{def}}{=} P(\text{k-string of H's appears among } m \text{ flips})}$$

so that, for us,

$$\boxed{\begin{aligned} P_n &= P(E) \\ P_{n-j} &= P(E \mid F_j) \end{aligned}}$$

$\leftarrow$ since the k-string must appear among the remaining $n-j$ flips

and the multiplication principle becomes

$$\begin{aligned}
P_n &= \underbrace{P(E \mid H_1)}_{=1} P(H_1) + \sum_{j=1}^{k} \underbrace{P(E \mid F_j)}_{=P_{n-j}} P(F_j) \\
&= \underbrace{P(H_1)}_{=p^k} + \sum_{j=1}^{k} P_{n-j} \cdot \underbrace{P(F_j)}_{=p^{j-1}(1-p)} \\
&= p^k + \sum_{j=1}^{k} P_{n-j} \cdot p^{j-1}(1-p)
\end{aligned}$$

Since for $m = 1, \ldots, k-1$ there are no strings of length k,

$$\boxed{P_1 = P_2 = \cdots = P_{k-1} = 0}$$

while

$$\boxed{P_k = p^k}$$

we observe a recursion relation:

$$P_{k+1} = p^k + \sum_{j=1}^{k} P_{(k+1)-j}\, p^{j-1}(1-p)$$

$$= p^k + \underbrace{P_k}_{=p^k} \cdot p^{1-1}(1-p)$$

$$+ \underbrace{P_{k-1} p^{2-1}(1-p) + \cdots + P_1 p^{k-1}(1-p)}_{=0}$$

$$= p^k + p^k(1-p)$$

$$P_{k+2} = p^k + \sum_{j=1}^{k} P_{(k+2)-j}\, p^{j-1}(1-p)$$

$$= p^k + \underbrace{P_{k+1}}_{=p^k + p^k(1-p)} p^0(1-p) + \underbrace{P_k}_{=p^k} \cdot p^1(1-p) + \underbrace{\sum_{j=3}^{k} P_{k+2-j} \cdot p^{j-1}(1-p)}_{=0}$$

$$= p^k + p^k(1-p) + p^k(1-p)^2 + p^{k+1}(1-p)$$

$$P_{k+3} = p^k + \sum_{j=1}^{k} P_{(k+3)-j} \cdot p^{j-1}(1-p)$$

$$= p^k + P_{k+2} \cdot p^0(1-p) + P_{k+1}\, p^1(1-p) + P_k\, p^2(1-p)$$

$$= p^k + \left[p^k\left(1+(1-p)+(1-p)^2\right) + p^{k+1}(1-p)\right](1-p)$$

$$+ \left[p^k\left(1+(1-p)\right)\right] p(1-p) + p^k \cdot p^2(1-p)$$

$$= p^k\left(1+(1-p)+(1-p)^2+(1-p)^3\right)$$

$$+ p^{k+1}(1-p)^2 + p^{k+1}\left((1-p)+(1-p)^2\right) + p^{k+2}(1-p)$$

$$= p^k \sum_{l=0}^{3}(1-p)^l + p^{k+1}(1-p)\left(1+2(1-p)\right) + p^{k+2}(1-p)$$

etc.

Exercise: Suppose we take $n = 7$, $k = 5$, $p = 2/3$. What is the probability that we get a sequence of 3 heads in 5 tosses of a coin with $p = P(H) = 2/3$?

<u>A2</u>: Let $\boxed{S = \{\tau_1\tau_2\cdots\tau_n \mid \tau_i \in \{H,T\},\ \forall i\}}$ be our sample space, and define the random variable

$$\boxed{\begin{aligned} &L_n : S \longrightarrow \mathbb{R} \\ &L_n(\tau_1\cdots\tau_n) = \text{max \# consecutive } H\text{'s in } \tau_1\cdots\tau_n \end{aligned}}$$

If E = there is a string of k H's, then

$$P(E) = P(L_n \geq k)$$

To compute this, define the events

$$\boxed{\begin{aligned} &E_i = \left\{ \begin{aligned} &\tau_i = \tau_{i+1} = \cdots = \tau_{i+k-1} = H \\ &\&\ \tau_{i+k} = T \end{aligned} \right\} \text{ for } i = 1, \ldots, n-k \\ \\ &E_{n-k+1} = \text{"}\tau_{n-k+1} = \cdots = \tau_n = H\text{"} \end{aligned}}$$

and note that

$$P(E) = P(L_n \geq k)$$
$$= P\left(\bigcup_{i=1}^{n-k+1} E_i\right)$$
$$= \sum_{r=1}^{n-k+1} (-1)^{r+1} \sum_{i_1 < \cdots < i_r} P(E_{i_1} \cap \cdots \cap E_{i_r})$$

Letting $S_i = \{$flip #'s i through $i+k$, $i = 1, \ldots, n-k+1\}$
$= \{i, i+1, \ldots, i+k \mid i = 1, \ldots, n-k+1\}$

we observe that

$$P(E_{i_1} \cap \cdots \cap E_{i_r}) = \begin{cases} 0, & \text{if } S_{i_j} \cap S_{i_l} \neq \emptyset \\ p^{rk}(1-p)^r, & S_{i_j} \cap S_{i_l} = \emptyset \text{ for all } i_j \neq i_l \text{ in } i_1, \ldots, i_r \end{cases}$$

$i_r < n-k+1$ here

$$\Rightarrow \sum_{i_1 < \cdots < i_r} P(E_{i_1} \cap \cdots \cap E_{i_r}) = \binom{n-rk}{r} p^{rk}(1-p)^r$$

$\binom{n-rk}{r}$ $\Leftarrow$ $r(k+1)$ slots occupied by S_{i_j}'s, $n - r(k+1)$ by the rest

= # positions of S_i among rest, & following each S_i w/ T and

$$P(E_{i_1} \cap \cdots \cap E_{i_{r-1}} \cap E_{n-k+1})$$

$$= \begin{cases} 0, \text{ if overlap} \\ (p^k(1-p))^{r-1} p^k = p^{rk}(1-p)^{r-1} \end{cases}$$

$$\Rightarrow \sum_{i_1 < \cdots < n-k+1} P(E_{i_1} \cap \cdots \cap E_{n-k+1}) = \binom{n-rk}{r-1} p^{rk}(1-p)^{r-1}$$

Therefore,

$$\boxed{\begin{aligned} P(E) &= P(L_n \geq k) \\ &= \sum_{r=1}^{n-k+1} (-1)^{r+1} \left[\binom{n-rk}{r} + \frac{1}{p}\binom{n-rk}{r-1} \right] p^{rk}(1-p)^r \end{aligned}}$$

<u>Exercise</u> Compute $P(E)$ for $n = 7$, $k = 5$, $p = 2/3$.

Let us now use a <u>Poisson approximation</u> to this problem, $P(\underbrace{\text{there is a string of } k\ H\text{'s}}_{E\,=})$.

<u>(failed) attempt #1</u>: Let H_i = event that the k flips $i, \ldots, i+k-1$ are all H, as before, & note again $P(H_i) = p^k$.

The problem here is the overlap of the H_i, for example

$$P(H_2 | H_1) = \frac{\overbrace{P(H_1 \cap H_2)}^{=\, p^{k+1}}}{\underbrace{P(H_2)}_{=\, p^k}}$$

$$= \frac{p^{k+1}}{p^k} = p$$

so that $P(H_1 \cap H_2) = p^{k+1} \neq p^{2k} = p^k p^k$
$= P(H_1)P(H_2)$

& H_1 & H_2 are not independent.
(i.e. $P(H_2|H_1) = p \neq p^k = P(H_2)$)

attempt #2: We go instead with the events

$i=1,\dots,n-k$ $\left\{ E_i = \text{flips } i,\dots,i+k-1 \text{ are all } H \text{ AND flip } i+k \text{ is } T \right.$

$i=n-k+1$ $\left\{ E_{n-k+1} = \text{flips } n-k+1,\dots,n \text{ are } H \right.$

so that, even if there is a string of H's longer than k, we consider only the last k terms of that sequence. This makes the events indep.:

(or close to it)

$$\boxed{P(E_i \mid E_j) = \begin{cases} P(E_i), & \text{if } E_i \,\&\, E_j \text{ overlap} \\ 0, & \text{if } E_i \,\&\, E_j \text{ do not overlap} \end{cases}}$$

for $i \neq j$, so that $P(E_i \cap E_j) = P(E_i)P(E_j)$ in the 1st case & $P(E_i \cap E_j) = 0 = 0 \cdot P(E_j)$ in the 2nd, $\approx P(E_i)P(E_j)$

for, in any case we require p^k small, &

$$\left.\begin{aligned} P(E_i) &= p^k(1-p), \quad i=1,\dots,n-k \\ P(E_{n-k+1}) &= p^k \end{aligned}\right\} \approx 0$$

Let $S = \{ \tau_1 \tau_2 \cdots \tau_n \mid \tau_i \in \{H, T\} \}$ be the sample space of outcomes of n flips of the coin, considered as strings of H & T, & let, as before,

$$E_i = \text{flips } \left.\begin{array}{l} \tau_i = \cdots = \tau_{i+k-1} = H \\ \& \quad \tau_{i+k} = T \end{array}\right\} \; i = 1, \ldots, n-k$$

$$E_{n-k+1} = \text{flips } \tau_{n-k+1} = \cdots = \tau_n = H$$

be events in $\mathcal{P}(S)$. Define

$$X : S \to \mathbb{R}, \quad X = \# \, E_i \text{ occurring in } \tau_1 \tau_2 \cdots \tau_n$$

so that

$$\begin{aligned} E[X] &= \sum_x x p(x) \\ &= \sum_{i=0}^{\lceil \frac{n}{2k} \rceil} i P(X = i) \\ &= 0 \cdot P(X=0) + 1 \cdot P(X=1) \\ &\quad + \cdots + \lceil n/2k \rceil P(X = \lceil n/2k \rceil) \end{aligned}$$

$$= \overbrace{\sum_{i=1}^{n-k+1} P(E_i)}^{X=1} + 2\overbrace{\sum_{i,j} P(E_i \cap E_j)}^{X=2 \begin{cases} p^{2k}(1-p) \approx 0 \ \text{or} = 0 \\ \text{or} \\ = p^{2k}(1-p)^2 \approx 0 \end{cases}}$$

$$+ \cdots + \left\lceil \frac{n}{2k} \right\rceil \underbrace{\sum_{i_1 < \cdots < i_{\lceil n/2k \rceil}} P(E_{i_1} \cap \cdots \cap E_{i_{\lceil n/2k \rceil}})}_{\approx 0}$$

$$\approx \sum_{i=1}^{n-k+1} P(E_i)$$

$$= \underbrace{(n-k)p^k(1-p) + p^k}_{E[X] = \lambda}$$

Therefore, to work out $P(E)$, define the random variable

$$\boxed{\begin{array}{l} L_n : S \to \mathbb{R} \\ L_n(\tau_1 \cdots \tau_n) = \text{max \# consecutive } H\text{'s in } \tau_1 \cdots \tau_n \end{array}}$$

and note that

$$\boxed{P(E) = P\left(\bigcup_{i=1}^{n-k+1} E_i\right) = P(L_n \geq k)}$$

Since

$$P(L_n < k) = P(X=0)$$

$$\approx e^{-\lambda} \cdot \frac{\lambda^0}{0!}$$

$$= e^{-\{(n-k)p^k(1-p)+p^k\}}$$

we have

$$\boxed{\begin{aligned} P(L_n \geq k) &\approx 1 - P(L_n < k) \\ &= 1 - e^{-\{(n-k)p^k(1-p)+p^k\}} \end{aligned}}$$

Let us illustrate this with the special case of $p = 1-p = \frac{1}{2}$, so that

$$e^{-\lambda} = e^{-\{(n-k)\cdot\frac{1}{2^{k+1}} + \frac{1}{2^k}\}}$$

$$= e^{-\{\frac{n}{2^{k+1}} + \underbrace{\frac{2-k}{2^{k+1}}}_{\approx 0 \text{ for large } k}\}}$$

$$\approx e^{-n/2^{k+1}}$$

$$= e^{-n/2^j \cdot 2^{i+1}} \qquad (k = i+j,$$

$$= e^{-1/2^{i+1}} \qquad j = \log_2 n\ )$$

$*$ n even

Then, since

$$P(L_n = i+j) = P(i+j \le L_n < i+j+1)$$

$$= P(L_n < i+j+1) - P(L_n < i+j)$$

$$= e^{-1/2^{i+2}} - e^{-1/2^{i+1}}$$

we have, for $j = \log_2 n$, n large,

$$P(L_n < j-3) \cong e^{-1/2^{-3+1}} = e^{-4} \approx 0.0183$$

L_n within 2 of $j-1$ has prob. their sum ≈ 0.86

$$P(L_n = j-3) \cong e^{-1/2^{-3+2}} - e^{-1/2^{-3+1}}$$

$$\cong e^{-2} - e^{-4} \approx 0.1170$$

$$P(L_n = j-2) \cong e^{-1/2^{-2+2}} - e^{-1/2^{-2+1}}$$

$$\cong e^{-1} - e^{-2} \approx 0.2325$$

$$P(L_n = j-1) \cong e^{-1/2} - e^{-1} \approx 0.2387$$

$$P(L_n = j) \cong e^{-1/4} - e^{-1/2} \approx 0.1723$$

$$P(L_n = j+1) \cong e^{-1/8} - e^{-1/4} \approx 0.1037$$

$$P(L_n = j+2) \approx e^{-1/16} - e^{-1/8} \approx 0.0569$$

$$\vdots$$