L1

Longest Run of Heads.

HW1 posted on Friday. For credit. Due 1 wk later.

This lecture is based on: "The Longest Run of Heads" by Schilling (1990). You should read it: it is well-written and very manageable.

Distribution of the longest run of heads in 3 (fair) coin tosses.

Sample space:

$\underset{3}{HHH}$, $\underset{2}{HHT}$, $\underset{1}{HTH}$, $\underset{1}{HTT}$,
$\underset{2}{THH}$, $\underset{1}{THT}$, $\underset{1}{TTH}$, $\underset{0}{TTT}$.

length of longest head run, x	probability, $P(R_3=x)$
0	1/8
1	4/8
2	2/8
3	1/8.

$$\Rightarrow \mathbb{E}[R_3] = 0\,\frac{1}{8} + 1\cdot\frac{4}{8} + 2\cdot\frac{2}{8} + 3\cdot\frac{1}{8}$$

$$= \frac{4+4+3}{8} = \frac{11}{8} = 1\,{}^{3}/_{8}.$$

Hmm, even in a sequence as short as 3 flips, you'd expect more than 1 heads in a row! But people who haven't done the math tend to underestimate this number.

When n becomes as large as 6, it becomes quite laborious to compute the distribution by enumerating the sample space.

Recursive approach

Let
$$F_n(x) = P(R_n \le x)$$

Let $A_n(x)$ be the # sequences of length n in which the longest run of heads has length $\le x$. Then
$$F_n(x) = \frac{A_n(x)}{2^n}.$$

But how do we compute $A_n(x)$?

$x = 3$: favorable outcomes	R_n
... TT ...	0
... THT ...	1
... THHT ...	2
... THHHT ...	3

$n \leq 3$: every possible outcome is favorable! (Review sample space of earlier example.) $\Rightarrow$ $A_n(3) = 2^n$ $(n \leq 3)$

$n > 3$?

The key is to partition the set of favorable outcomes (sequences) according to the number of heads before the first tail:

T . . .

HT . . .

HHT . . .

HHHT . . .

$\boxed{\text{WE STOP HERE!}}$

→ eg. HHHHT . . . has $R_n > 3$

Note, also, that each ". . ." is a sequence of length $m < n$ s.t. $R_m \leq 3$. Another way of saying this is the set of seqs s.t. $R_n \leq 3$ is made up of the sets:

TS_{n-1}
HTS_{n-2}
$HHTS_{n-3}$
$HHHTS_{n-4}$

where S_m is a sequence of length m that contains a run of 3 heads or fewer.

Thus:

$$A_n(3) = \#(TS_{n-1}) + \#(HTS_{n-2}) + \#(HHTS_{n-2}) + \#(HHHTS_{n-4})$$

$$= \#(S_{n-1}) + \#(S_{n-2}) + \#(S_{n-2}) + \#(S_{n-4})$$

$$\Rightarrow A_n(3) = A_{n-1}(3) + A_{n-2}(3) + A_{n-3}(3) + A_{n-4}(3) \quad (*)$$

Now: $A_0(3) = 2^0 = 1$

$A_1(3) = 2^1 = 2$

$A_2(3) = 2^2 = 4$

$A_3(3) = 2^3 = 8$

$(*) \Rightarrow$

$$A_4(3) = A_3(3) + A_2(3) + A_1(3) + A_0(3)$$

$$= 8 + 4 + 2 + 1.$$

$$= 15$$

$$A_5(3) = A_4(3) + A_3(3) + A_2(3) + A_1(3)$$

$$= 15 + 8 + 4 + 2$$

$$= 29$$

Similarly:

$$A_6(3) = 56$$

$$A_7(3) = 108$$

$$A_8(3) = 208$$

How can we use this to predict whether a given sequence was generated by a truly random process (a fair coin), or not?

Answer: compute the probability that the sequence contains a run of x heads, say.

Recall that

$$P(R_n \le x) = \frac{A_n(x)}{2^n}$$

We want $P(R_n = x)$. We find it by observing that:

$$F_n(x) = P(R_n \le x) = \sum_{y=0}^{x} P(R_n = y)$$

$$\Rightarrow \quad F_n(x) - F_n(x-1) = P(R_n = x)$$

Thus
$$P(R_n = x) = \frac{A_n(x) - A_n(x-1)}{2^n}$$

So if we want $P(R_n = 3)$, then we need not only $A_n(3)$, but also $A_n(2)$. Clearly, we need an algorithm for computing $A_n(x)$ for arbitrary n and x. If you review our derivation of $A_n(x)$ for $x=3$, you'll see that the formula we obtained generalizes to arbitrary x in an obvious manner:

$$A_n(x) = \begin{cases} \sum_{j=0}^{x} A_{n-1-j}(x) & \text{for } n > x \\ 2^n & n \leq x \end{cases}$$

Figure 3 illustrates $P(R_n = x)$ for various n. The remarkable fact is that the distribution just drifts to the right w/o broadening. This means that you can predict the value of R_n to within 3 heads w/ high confidence for any n, no matter how big! In other words, sequences whose longest run has a length outside this "confidence interval" are likely to be fake!

log n law

View each head run in a coin tossing seq as beginning w/ the first toss or immediately after tails occurs. (We are allowing runs of length zero.)

		R
examples:	T T T T	0
	T H T	1
	T H H T T	2
	T H H H T	3

Let prob of heads = p. Let q = 1 - p.
In a seq of length n there will be $\simeq nq$ runs in all, since this is the expected # tails.

Of these runs, a fraction p will start w/ a head and therefore contain @ least one head.

Similarly, of all runs, a fraction p^2 will start w/ start w/ 2 heads, and therefore contain @ least 2 heads.

In general, the expected number of sequences containing @ least x heads is: nqp^x.

Now, it is quite possible that $nqp^x \gg 1$. When this is the case, we can expect more than one run of length x or greater, ie. $P(R_n \geq x) \simeq 1$

For larger x st. $nqp^x \ll 1$, we expect less than one such run: $P(R_n \geq x) \ll 1$.

Only when $nqp^x \simeq 1$ do we expect exactly one run of length $\geq x$, ie. $P(R_n \geq x) \sim 0.5$. The value of x at which $P(R_n \geq x)$ crosses over from $\ll 1$ to exponentially to $\underline{1}$ is $\simeq \mathbb{E}[R_n]$, ie.

$$nqp^{\mathbb{E}[R_n]} \simeq 1$$

$$\Rightarrow \quad nq \simeq \left(\frac{1}{p}\right)^{\mathbb{E}[R_n]}$$

$$\Rightarrow \quad \boxed{\mathbb{E}[R_n] \simeq \log_{1/p}(nq)}$$

eg $p = q = 1/2 \Rightarrow \mathbb{E}[R_n] \simeq \log_2(n/2)$

n	50	100	200
$E[R_n]$	4.6	5.6	6.6