ECS315 2015/1 Part I.3 Dr.Prapun

4.38. Further reading on combinatorial ideas: the pigeon-hole principle, inclusion-exclusion principle, generating functions and recurrence relations, and flows in networks.

4.4 Famous Example: Galileo and the Duke of Tuscany

Example 4.39. When you toss three dice, the chance of the sum being 10 is greater than the chance of the sum being 9.

- The Grand Duke of Tuscany "ordered" Galileo to explain a paradox arising in the experiment of tossing three dice [2]:

  > "Why, although there were an equal number of 6 partitions of the numbers 9 and 10, did experience state that the chance of throwing a total 9 with three fair dice was less than that of throwing a total of 10?"

- Partitions of sums 11, 12, 9 and 10 of the game of three fair dice:

1+4+6=11	1+5+6=12	3+3+3=9	1+3+6=10
2+3+6=11	2+4+6=12	1+2+6=9	1+4+5=10
2+4+5=11	3+4+5=12	1+3+5=9	2+2+6=10
1+5+5=11	2+5+5=12	1+4+4=9	2+3+5=10
3+3+5=11	3+3+6=12	2+2+5=9	2+4+4=10
3+4+4=11	4+4+4=12	2+3+4=9	3+3+3=10

  The partitions above are not equivalent. For example, from the addenda 1, 2, 6, the sum 9 can come up in $3! = 6$ different ways; from the addenda 2, 2, 5, the sum 9 can come up in $\frac{3!}{2!1!} = 3$ different ways; the sum 9 can come up in only one way from 3, 3, 3.

- ***Remarks***: Let X_i be the outcome of the ith dice and S_n be the sum $X_1 + X_2 + \cdots + X_n$.

  (a) $P[S_3 = 9] = P[S_3 = 12] = \frac{25}{6^3} < \frac{27}{6^3} = P[S_3 = 10] = P[S_3 = 11]$. Note that the difference between the two probabilities is only $\frac{1}{108}$.

  (b) The range of S_n is from n to $6n$. So, there are $6n-n+1 = 5n+1$ possible values.

  (c) The pmf of S_n is symmetric around its expected value at $\frac{n+6n}{2} = \frac{7n}{2}$.

  ◦ $P[S_n = m] = P[S_n = 7n - m]$.

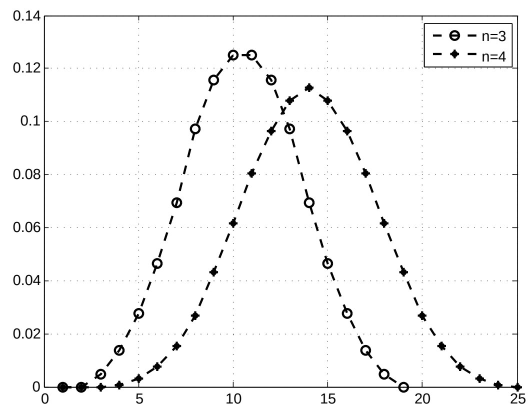

Figure 2: pmf of S_n for $n = 3$ and $n = 4$.

4.5 Application: Success Runs

Example 4.40. We are all familiar with "success runs" in many different contexts. For example, we may be or follow a tennis player and count the number of consecutive times the player's first serve is good. Or we may consider a run of forehand winners. A basketball player may be on a "**hot streak**" and hit his or her shots perfectly for a number of plays in row.

In all the examples, whether you should or should not be amazed by the observation depends on a lot of other information. There may be perfectly reasonable explanations for any particular success run. But we should be curious as to whether randomness could also be a perfectly reasonable explanation. Could the hot streak of a player simply be a snapshot of a random process, one that we particularly like and therefore pay attention to?

In 1985, cognitive psychologists Amos Taversky and Thomas Gilovich examined[12] the shooting performance of the Philadelphia 76ers, Boston Celtics and Cornell University's men's basketball team. They sought to discover whether a player's previous shot had any predictive effect on his or her next shot. Despite basketball fans' and players' widespread belief in hot streaks, the researchers found no support for the concept. (No evidence of nonrandom behavior.) [14, p 178]

4.41. Academics call the mistaken impression that a random streak is due to extraordinary performance the **hot-hand fallacy**. Much of the work on the hot-hand fallacy has been done in the context of sports because in sports, performance is easy to define and measure. Also, the rules of the game are clear and definite, data are plentiful and public, and situations of interest are replicated repeatedly. Not to mention that the subject gives academics a way to attend games and pretend they are working. [14, p 178]

Example 4.42. Suppose that two people are separately asked to toss a fair coin 120 times and take note of the results. Heads is noted as a "one" and tails as a "zero". The following two lists of compiled zeros and ones result

1 1 0 0 1 0 0 1 0 1 1 0 0 1 0 0 0 1 1 0 1 0 1 0 0 1 1 0 1 0
0 1 0 1 0 1 1 0 1 1 0 0 1 1 0 1 1 1 0 1 0 0 1 0 0 1 1 0 1 0
0 1 1 0 1 0 0 1 1 0 1 0 1 1 0 0 1 1 1 0 0 1 0 1 0 1 0 0 0 1
0 1 0 1 0 1 0 1 0 1 1 0 0 1 0 0 1 0 1 1 0 0 1 0 0 1 1 0 1 1

and

[12] "The Hot Hand in Basketball: On the Misperception of Random Sequences"

1 1 1 0 0 0 1 1 1 0 1 0 1 1 1 1 1 1 0 1 0 0 0 1 1 0 0 1 1 0
1 0 1 0 0 0 1 1 0 1 0 0 1 1 1 0 1 0 0 0 0 1 0 1 1 1 0 1 1 0
0 1 1 1 0 1 1 0 0 1 1 1 1 1 1 0 1 1 0 1 0 1 1 1 0 0 0 0 0 0
0 0 1 1 0 1 1 1 0 1 1 1 1 0 1 1 1 1 0 1 0 1 1 0 1 1 0 1 0 1

One of the two individuals has cheated and has fabricated a list of numbers without having tossed the coin. Which list is more likely be the fabricated list? [21, Ex. 7.1 p 42–43]

The answer is later provided in Example 4.48.

Definition 4.43. A **run** is a sequence of more than one consecutive identical outcomes, also known as a **clump**.

Definition 4.44. Let R_n represent the length of the longest run of heads in n independent tosses of a fair coin. Let $\mathcal{A}_n(x)$ be the set of (head/tail) sequences of length n in which the longest run of heads does not exceed x. Let $a_n(x) = \|\mathcal{A}_n(x)\|$.

Example 4.45. If a fair coin is flipped, say, three times, we can easily list all possible sequences:

HHH, HHT, HTH, HTT, THH, THT, TTH, TTT

and accordingly derive:

x	$P[R_3 = x]$	$a_3(x)$
0	1/8	1
1	4/8	4
2	2/8	7
3	1/8	8

4.46. Consider $a_n(x)$. Note that if $n \le x$, then $a_n(x) = 2^n$ because any outcome is a favorable one. (It is impossible to get more than three heads in three coin tosses). For $n > x$, we can partition $\mathcal{A}_n(x)$ by the position k of the first tail. Observe that k must be $\le x+1$ otherwise we will have more than x consecutive heads in the sequence which contradicts the definition of $\mathcal{A}_n(x)$. For each $k \in \{1, 2, \dots, x+1\}$, the favorable sequences are in the form

$$\underbrace{\text{HH}\dots\text{H}}_{k-1 \text{ heads}}\text{T}\ \underbrace{\text{XX}\dots\text{X}}_{n-k \text{ positions}}$$

where, to keep the sequences in $\mathcal{A}_n(x)$, the last $n-k$ positions[13] must be in $\mathcal{A}_{n-k}(x)$. Thus,

$$a_n(x) = \sum_{k=1}^{x+1} a_{n-k}(x) \text{ for } n > x.$$

In conclusion, we have

$$a_n(x) = \begin{cases} \sum_{j=0}^{x} a_{n-j-1}(x), & n > x, \\ 2^n & n \leq x \end{cases}$$

[20]. The following MATLAB function calculates $a_n(x)$

```
function a = a_nx(n,x)
a = [2.^(1:x) zeros(1,n-x)];
a(x+1) = 1+sum(a(1:x));
for k = (x+2):n
    a(k) = sum(a((k-1-x):(k-1)));
end
a = a(n);
```

4.47. Similar technique can be used to construct $\mathcal{B}_n(x)$ defined as the set of sequences of length n in which the longest run of heads and the longest run of tails do not exceed x. To check whether a sequence is in $\mathcal{B}_n(x)$, first we convert it into sequence of S and D by checking each adjacent pair of coin tosses in the original sequence. S means the pair have same outcome and D means they are different. This process gives a sequence of length $n-1$. Observe that a string of $x-1$ consecutive S's is equivalent to a run of length x. This put us back to the earlier problem of finding $a_n(x)$ where the roles of H and T are now played by S and D, respectively. (The length of the sequence changes from n to $n-1$ and the max run length is $x-1$ for S instead of x for H.) Hence, $b_n(x) = \|\mathcal{B}_n(x)\|$ can be found by

$$b_n(x) = 2a_{n-1}(x-1)$$

[20].

[13]Strictly speaking, we need to consider the case when $n = x+1$ separately. In such case, when $k = x+1$, we have $\mathcal{A}_0(x)$. This is because the sequence starts with x heads, then a tail, and no more space left. In which case, this part of the partition has only one element; so we should define $a_0(x) = 1$. Fortunately, for $x \geq 1$, this is automatically satisfied in $a_n(x) = 2^n$.

Example 4.48. Continue from Example 4.42. We can check that in 120 tosses of a fair coin, there is a very large probability that at some point during the tossing process, a sequence of five or more heads or five or more tails will naturally occur. The probability of this is

$$\frac{2^{120} - b_{120}(4)}{2^{120}} \approx 0.9865.$$

0.9865. In contrast to the second list, the first list shows no such sequence of five heads in a row or five tails in a row. In the first list, the longest sequence of either heads or tails consists of three in a row. In 120 tosses of a fair coin, the probability of the longest sequence consisting of three or less in a row is equal to

$$\frac{b_{120}(3)}{2^{120}} \approx 0.000053,$$

which is extremely small indeed. Thus, the first list is almost certainly a fake. Most people tend to avoid noting long sequences of consecutive heads or tails. Truly random sequences do not share this human tendency! [21, Ex. 7.1 p 42–43]