A PROBLEM IN THE THEORY OF PROBABILITY

By T. V. Narayana
Institut Henri Poincaré, Paris

Statement of Problem

Suppose that we are given a coin for which the probability is p for obtaining heads and consequently $q = 1 - p$ for obtaining tails, where $p > q$. In a long series of independent trials with this coin, we may thus expect with probability 1 that the total number of heads obtained will exceed the total number of tails obtained, since $p > q$. Let us now consider a game g_1 to be played as follows: we shall stop our series of trials at the first trial when the accumulated number of heads exceeds the accumulated number of tails by exactly *one*. The game g_1 can thus end only at the $(2n + 1)$-st trial, $n = 0, 1, 2, \ldots$ and we shall have n tails and $(n + 1)$ heads in such a sequence. Obviously the last trial in such a sequence is a head and for $n > 0$, the first trial is necessarily a tail. For the first few values of n, we obtain the following sequences in the game g_1, where we denote a head by x and a tail by o:

$$x;\ oxx;\ ooxxx,\ oxoxx;\ oxooxxx,\ oooxxxx,$$
$$ooxxoxx,\ ooxoxxx,\ oxoxoxx;\ \text{etc.}$$

If f_n is the probability that the game g_1 ends at the n-th trial, evidently $f_n = 0$ whenever n is even.

Recurrent Events and the Game g_1

Let us consider for $n > 0$, a sequence of $(2n + 1)$ trials assuming that the game g_1 ends at the $(2n + 1)$-st trial. If we delete the final x from such a sequence, we have a sequence of $2n$ x's and o's, where the first trial is a o and at the $2n$-th trial the accumulated number of x's and o's are equal, being n each; further, if the accumulated number of x's and o's had been equal at the $(2k)$-th trial, $(k < n)$, the $(2k + 1)$-st trial is always a o. In other words, we have a "return to equilibrium" (in coin tossing) at the $(2n)$-th trial and the $(2n + 1)$-st trial is a x, *with the proviso that if there had been a return to equilibrium at the* (2k)-*th trial* (k < n), *the* (2k + 1)-*st trial is a* o. From the theory of recurrent events due to Feller[1, 2]* we obtain easily the

* Small numerals indicate the references listed at the end of the paper.

probability, f_{2n+1}, that the game g_1 ends at the $(2n+1)$-st trial, $n = 0, 1, 2, \ldots$ as

$$f_{2n+1} = \frac{1}{n+1} \binom{2n}{n} p^{n+1} q^n.$$

We obtain also the generating function of the probability distribution determined by g_1 as

$$g_1(s) = \sum_{n=0}^{\infty} f_n s^n$$

$$= \sum_{n=0}^{\infty} \frac{1}{n+1} \binom{2n}{n} p^{n+1} q^n s^{2n+1}$$

$$= \frac{1-(1-4pqs^2)^{\frac{1}{2}}}{2qs}.$$

Denoting by $E(g_1)$ and $V(g_1)$ the expected number and the variance of the number of trials before g_1 ends, we see by differentiation of the generating function, that

$$E(g_1) = \frac{1}{p-q},$$

$$V(g_1) = \frac{4qp}{(p-q)^3}.$$

The Game g_2

We consider now the game g_2, where we shall stop our series of trials at the *first* trial when the total number of heads exceeds the total number of tails by exactly 2. We shall assume, as before, that $p > q$, where p and $q = 1 - p$ are the probabilities of obtaining heads and tails respectively with our coin. We remark that the game g_2 can end only at the $(2n+2)$-nd trial $n = 0, 1, 2, \ldots$ and that the last two trials are necessarily heads (x's). For the first three values of n, we have the following sequences in the game g_2:

xx; *oxxx*, *xoxx*; *ooxxxx*, *xooxxx*, *xoxoxx*,
oxxoxx, *oxoxxx*.

We observe the important fact that any sequence of g_2 of length say $(2n+2)$, may be obtained by juxtaposition of two sequences of g_1 of lengths $(2n_1+1)$, $(2n_2+1)$, where $n_1 + n_2 = n$ and $n_1 \geq 0$,

$n_2 \geqslant 0$ are integers. Thus by juxtaposing any two sequences whatever of g_1 we obtain a sequence of g_2. Conversely, any sequence of g_2 can be split up into two g_1 sequences. This is obvious, since given a sequence of g_2, we know that by proceeding from the first trial to the last, *i.e.*, $(2n+2)$-nd trial, we end at the *first* trial where the number of heads exceeds by exactly 2 the number of tails. *A fortiori*, we shall have to obtain at some point, the first trial, trial $(2n_1+1)$ say, $n_1 \geqslant 0$, where the number of heads exceeds by exactly 1 the number of tails. Thus the first $(2n_1+1)$ trials constitute a sequence belonging to g_1 and so do the remaining trials from $(2n_1+2), \ldots (2n+2)$. Hence, the generating function of g_2 will be given by

$$g_2(s) = [g_1(s)]^2 = \frac{1-(1-4pqs^2)^{\frac{1}{2}}}{2q^2s^2} - \frac{p}{q}$$

$$= \sum_{n=0}^{\infty} \frac{2}{n+2} \binom{2n+1}{n} p^{n+2} q^n s^{2n+2}.$$

Setting $X = g_1(s)$, we have

$$X^2 = \frac{X}{qs} - \frac{p}{q},$$

or

$$qsX^2 - X + ps = 0,$$

X being, in fact, the negative root of the quadratic.

Hence,

$$qsX^2 = X - ps \tag{A}$$

Generating Function of the Game g_r

Generally, let $g_r(s)$ denote the generating function of the game where we stop our series of trials at the *first* trial when the accumulated number of heads exceeds the accumulated number of tails by exactly r, $r > 0$ being any integer. Then, by a similar reasoning

$$g_r(s) = [g_1(s)]^r = X^r.$$

For example,

$$g_3(s) = [g_1(s)]^3$$
$$= [g_1(s)] \cdot [g_2(s)],$$

and a simple calculation leads us to

$$g_3(s) = \sum_{n=0}^{\infty} \frac{3}{n+3} \binom{2n+2}{n} p^{n+3} q^n s^{2n+3}.$$

We shall prove now by induction that, for all integral $r > 0$

$$g_r(s) = \sum_{n=0}^{\infty} \frac{r}{n+r} \binom{2n+r-1}{n} p^{n+r} q^n s^{2n+r}.$$

This relation being true for $r = 1, 2, 3$ let us assume that it is true for all values of r up to and including a given r. We have now to demonstrate that

$$g_{r+1}(s) = \sum_{n=0}^{\infty} \frac{r+1}{n+r+1} \binom{2n+r}{n} p^{n+r+1} q^n s^{2n+r+1}.$$

Multiplying both sides of equation (A) by X^{n-1}, we obtain

$$qsX^{n+1} = X^n - psX^{n-1}$$

i.e.,

$$qsg_{r+1}(s) = g_r(s) - psg_{r-1}(s).\dagger$$

Thus,

$$qsg_{r+1}(s) = \sum_{n=0}^{\infty} \frac{r}{n+r} \binom{2n+r-1}{n} p^{n+r} q^n s^{2n+r}$$

$$- ps \sum_{n=0}^{\infty} \frac{r-1}{n+r-1} \binom{2n+r-2}{n} p^{n+r-1} q^n s^{2n+r-1}$$

$$= \sum_{n=1}^{\infty} \left\{ \frac{r}{n+r} \binom{2n+r-1}{n} - \frac{r-1}{n+r-1} \binom{2n+r-2}{n} \right\} p^{n+r} q^n s^{2n+r}.$$

† This result also follows from the following consideration:

Assuming $P(2n + r, r)$ to be the probability of A winning the game for the first time after r more successes than B's, it is obvious that

$$P(2n + r, r) = pP(2n + r - 1, r - 1) + qP(2n + r - 1, r + 1).$$

Writing this equation in terms of the generating function $g(s)$, it reduces to

$$g_r(s) = psg_{r-1}(s) + qsg_{r+1}(s).$$

Since $g_0(s)$ is known from Feller's work referred to by the author, the general solution for $g_r(s)$ can be obtained by the usual procedure followed for solving differential equations of the second order.—*Editor*.

Since

$$\frac{r}{n+r}\binom{2n+r-1}{n} - \frac{r-1}{n+r-1}\binom{2n+r-2}{n}$$
$$= \frac{r+1}{n+r}\binom{2n+r-2}{n-1}$$

we have

$$g_{r+1}(s) = \sum_{n=1}^{\infty} \frac{r+1}{n+r}\binom{2n+r-2}{n-1} p^{n+r} q^{n-1} s^{2n+r-1}$$
$$= \sum_{n=0}^{\infty} \frac{r+1}{n+r+1}\binom{2n+r}{n} p^{n+r+1} q^n s^{2n+r+1}.$$

THE "PROBLÈME DU SCRUTIN"

As an application of this result, we shall give a solution of the "problème du scrutin" worked out as an example by Henri Poincaré.[3]‡ The "problème du scrutin" is as follows: two candidates A and B stand for an election; a well-informed observer knows beforehand that A will obtain m votes and B n votes, where $m > n$. What is the probability that A holds the majority all the time during the scrutiny of the votes?

Consider a sequence, say S, of $(2n+r)$ observations which belongs to the game g_r. In such a sequence there are $n+r$ x's and n o's and the last observation of S, *i.e.*, the $(2n+r)$-th observation of S represents the *first* trial at which the number of x's is greater than the number of o's by exactly r. Let us consider the sequence S in reverse order and call it S', *i.e.*, the first trial of S is the $(2n+r)$-th trial of S', the second trial of S is the $(2n+r-1)$-th trial of S' ..., the $(2n+r)$-th trial of S is the first trial of S'. We may note also that S' read in inverse order will give us back again S. Let us make the convention that in S' an x represents a vote obtained by A and a o represents a vote obtained by B. Then a sequence S, *i.e.*, of $(2n+r)$ observations belonging to g_r gives rise to a sequence S' which represents one of the ways in which A getting $(n+r)$ votes retains the majority over B, who gets n votes.

For, evidently, the first vote of S' is an x and A holds the majority at the first observation. In order that A lose the majority,

‡ Professor G. Darmois kindly suggested to me to mention here that the solution of this problem is due to Désiré André. I express my very grateful thanks to him for kindly reading through the manuscript.

it is necessary at some stage that A and B have the same number of votes, say p each ($o < p \leqslant n$). If such were the case, *i.e.*, if A loses the majority at the $2p$-th observation in S', then the sequence S could not belong to g_r. This follows from the obvious fact, that in the remaining $(2n + r - 2p)$ votes of S, A must have a majority of r at least once, since A is certain to have it at the $(2n + r)$-th observation. Consequently, a subsequence of atmost $(2n + r - 2p)$ observations of S must have belonged to g_r. This is impossible since by assumption S belongs to g_r, and in S the $(2n + r)$-th trial is the *first* trial where A has a majority of r over B.

For the same reasons, every sequence S' in which A getting $(n + r)$ votes holds the majority over B getting n votes, when reversed, yields a sequence S of $(2n + r)$ observations belonging to g_r.

Hence in the "problème du scrutin", $r = m - n$ and

$$g_{m-n}(s) = \sum_{n=0}^{\infty} \frac{m-n}{m} \binom{m+n-1}{n} p^m q^n s^{m+n}$$

is the generating function of such sequences. The number of sequences, where A holds the majority over B, is thus

$$\frac{m-n}{m} \frac{(m+n-1)!}{n!\,(m-1)!}.$$

The total number of possible sequences being

$$\frac{(m+n)!}{m!\,n!},$$

the required probability is

$$\frac{m-n}{m+n}.$$

SOME IDENTITIES IN COMBINATORIAL ANALYSIS

Since obviously $g_r(1) = 1$, we have

$$\sum_{n=0}^{\infty} \frac{r}{n+r} \binom{2n+r-1}{n} p^{n+r} q^n = 1$$

for every integral $r > 0$, and $p > q$, $p + q = 1$. We observe, too, that for all integral r_1, r_2

$$g_{r_1}(s) \times g_{r_2}(s) = g_{r_1+r_2}(s)$$

i.e.,

$$\left\{\sum_{n=0}^{\infty}\frac{r_1}{n+r_1}\binom{2n+r_1-1}{n}p^{n+r_1}q^n s^{2n+r_1}\right\}\left\{\sum_{n=0}^{\infty}\frac{r_2}{n+r_2}\binom{2n+r_2-1}{n}p^{n+r_2}q^n s^{2n+r_2}\right\}$$

$$\equiv \sum_{n=0}^{\infty}\frac{r_1+r_2}{n+r_1+r_2}p^{n+r_1+r_2}q^n s^{2n+r_1+r_2}\binom{2n+r_1+r_2-1}{n}.$$

By equating coefficients of like powers of s we obtain from this identity the combinatorial formula:

$$\sum_{j=0}^{n}\frac{r_1}{n-j+r_1}\binom{2n-2j+r_1-1}{n-j}\frac{r_2}{j+r_2}\binom{2j+r_2-1}{j} = \frac{r_1+r_2}{n+r_1+r_2}\binom{2n+r_1+r_2-1}{n},$$

($r_1 > 0$, $r_2 > 0$, $n \geqslant 0$ being integers). Likewise, by differentiating the given identity and equating coefficients we obtain

$$\frac{r_1+r_2}{r_1 r_2}\binom{2n+r_1+r_2}{n} = \sum_{j=0}^{n}\left\{\binom{2j+r_1}{j}\frac{1}{n-j+r_2}\binom{2n-2j+r_2-1}{n-j} + \binom{2j+r_2}{j}\frac{1}{n-j+r_1}\binom{2n-2j+r_1-1}{n-j}\right\}.$$

This process can be continued indefinitely giving rise to a series of identities for positive integral r_1, r_2 (both being > 0). These identities may be interpreted as the relationships between the moments of the probability distributions defined by $g_{r_1}(s)$, $g_{r_2}(s)$, $g_{r_1+r_2}(s)$; $g_{r_1+r_2}(s)$ being obtained as a convolution of $g_{r_1}(s)$ and $g_{r_2}(s)$.

We can further generalise our identities by utilising the relation (for all integral $r_i > 0$, $i = 1, \ldots, k$)

$$g_{r_1+\cdots+r_k}(s) = \prod_{i=1}^{k} g_{r_i}(s)$$

i.e.,

$$\sum_{n=0}^{\infty}\frac{r_1+\ldots+r_k}{n+r_1+\ldots+r_k}\binom{2n+r_1+\ldots+r_k-1}{n}p^{n+r_1+\cdots+r_k}q^n s^{2n+r_1+\cdots+r_k}$$

$$\equiv \prod_{i=1}^{k}\left\{\sum_{n=0}^{\infty}\frac{r_i}{n+r_i}\binom{2n+r_i-1}{n}p^{n+r_i}q^n s^{2n+r_i}\right\}.$$

By taking the m-th derivatives *w.r.t.*'s on both sides of this identity and equating coefficients, we shall obtain a series of identities for $m = 0, 1, 2, \ldots$

SUMMARY

We consider in this paper a series of games g_r for $r=1, 2, \ldots$

For $r = 0$, the game g_r represents the "recurrent events" of Feller.

The generating function of g_r has been derived and the recurrence relation connecting the generating functions of g_{r-2}, g_{r-1} and g_r has been obtained.

An application to the André-Poincaré "problème du scrutin" is discussed as well as various combinatorial identities derivable from the generating functions.

REFERENCES

1. Feller, W. .. *An Introduction to Probability Theory and Its Applications*, John Wiley and Sons, Inc., New York, 1950.

2. Borel, E. .. *Traité du Calcul des Probabilités et ses Applications*, Paris, Gauthier Villars, 1925.

3. Poincaré, H. .. *Calcul des Probabilités*, Paris, Gauthier-Villars, 1912.