RESEARCH NOTES

Use of Complex Markoff's Chain in Testing Randomness

By Baikunth Nath Singh, M.Sc.
Banaras Hindu University

It is intended to consider in this paper a criterion for testing the randomness of a sequence of points, arranged on a line, which can assume one of the two different characters and when between the points a dependence of the type of a complex Markoff's chain is involved. The type of dependence we consider here is that the probability of the occurrence of the i-th point depends upon the knowledge of the occurrence of the preceding two only, *i.e.*, the $(i-2)$-th and $(i-1)$-th, no matter whatever information we may possess about the materialisation of the other preceding points.

We consider the case of free sampling, that is, when the character of each point is determined on the null hypothesis independently of the character of other points. Suppose there are m points on the line and the probability that a given point will assume one character (say black) or the other (say white) is p or q respectively subject to the condition that $p+q=1$. Since in this type of dependence three points are involved at a time we would contemplate the formation of blocks of three adjacent points in succession till all the points are exhausted. Thus with four points x x x x we have two blocks of three. We would further consider in each block (of three) all the possible differences, taken in the same sense, between the points and regard only those differences as 'successes' (or contributing to the moment) in which p happens to come before q and not *vice versa*. Thus in the arrangements *ppq*, *ppqpq*, *qppp* we get the contributions as 2, 4 and 0 respectively where, to be more specific, the differences between the first and the second; the first and the third; the second and the third points have been considered and the contributions from all the possible blocks of three neighbouring points have been pooled up. That is, we are considering the distribution of the total number of successes in blocks of three from m points possessing two characters. This total will give a method for determining whether a given sequence involving two characters can be taken to be a random one or not. This will also provide a test criterion for testing the difference between two samples.

We proceed to obtain the difference equation for the above distribution. Let $P_{ii'}(m, r)$ stand for the probability of obtaining r such 'successes' or differences out of m points where i, i' designate the character of the first two points of the sequence under consideration and i, i' assume the values either 1 or 2 according as the point is black or white. We then have

$$\begin{aligned}
P_{11}(m, r) &= pP_{11}(m-1, r) + pP_{12}(m-1, r-2)\\
P_{12}(m, r) &= pP_{21}(m-1, r-1) + pP_{22}(m-1, r-2)\\
P_{21}(m, r) &= qP_{11}(m-1, r) + qP_{12}(m-1, r-1)\\
P_{22}(m, r) &= qP_{21}(m-1, r) + qP_{22}(m-1, r)
\end{aligned}$$

In terms of the probability generating function ϕ, the above equations reduce to

$$\begin{aligned}
\phi_{11}(m) &= p\phi_{11}(m-1) + p\xi^2\phi_{12}(m-1)\\
\phi_{12}(m) &= p\xi\phi_{21}(m-1) + p\xi^2\phi_{22}(m-1)\\
\phi_{21}(m) &= q\phi_{11}(m-1) + q\xi\phi_{12}(m-1)\\
\phi_{22}(m) &= q\phi_{21}(m-1) + q\phi_{22}(m-1)
\end{aligned}$$

If the operator E be defined by the relation $E\phi(m) = \phi(m+1)$ we get

$$\begin{aligned}
(E-p)\phi_{11}(m-1) - p\xi^2\phi_{12}(m-1) &= 0\\
E\phi_{12}(m-1) - p\xi\phi_{21}(m-1) - p\xi^2\phi_{22}(m-1) &= 0\\
-q\phi_{11}(m-1) - q\xi\phi_{12}(m-1) + E\phi_{21}(m-1) &= 0\\
-q\phi_{21}(m-1) + (E-q)\phi_{22}(m-1) &= 0
\end{aligned}$$

Eliminating the ϕ's we get the determinantal equation

$$\begin{vmatrix} E-p & -p\xi^2 & 0 & 0\\ 0 & E & -p\xi & -p\xi^2\\ -q & -q\xi & E & 0\\ 0 & 0 & -q & E-q \end{vmatrix} \phi(m-1) = 0$$

The expansion and a little simplification of the above yields the difference equation as

$$\begin{aligned}
&\phi(m+3) - \phi(m+2) + pq(1-\xi^2)\phi(m+1)\\
&\quad + pq\xi^2(1-\xi)\phi(m) - p^2q^2\xi^2(1-\xi)^2\phi(m-1)\\
&\quad = 0.
\end{aligned}$$

It can be shown[1] that the above distribution will, for large values of m and providing p is not very small, tend to the normal form and in which case the standardised variate can be employed for testing the randomness.

We obtain now the mean and the variance of the above distribution which would be required in applying the test of significance. Advantage would be taken of the method of getting the factorial moments.[2] It can be seen that

$$\frac{\mu'_{[1]}}{1!} = (m-2)\{k_1'p^2q + k_1''pq^2\} \tag{I}$$

and

$$\begin{aligned}\frac{\mu'_{[2]}}{2!} &= (m-2)\{k_3'p^2q + k_3''pq^2\} \\ &+ (m-3)\{k_4'p^3q + k_4''' p^2q^2 + k_4''pq^3\} \\ &+ (m-4)\{k_5' p^4q + k_5''' p^3q^2 + k_5'''' p^2q^3 + k_5'' pq^4)\} \\ &+ \binom{m-4}{2}\{k_1'p^2q + k_1''pq^2\}^2\end{aligned} \tag{II}$$

where the k's are the constants to be determined.

It is to be noted that the constant coefficients, associated with the symmetrically placed probability expressions, *e.g.*, p^3q^2 and p^2q^3, remain the same as can be easily seen by taking the inverted arrangement of a particular permutation and changing p to q and q to p. Bearing this in mind the k's may be evaluated from the distribution of the total number of successes for the following compositions of three, four and five points:

(i) p^2q, (ii) p^3q, (iii) p^2q^2, (iv) p^4q, (v) p^3q^2.

Below is given the frequency distribution and the first and the second factorial sums ($S_{[1]}$ and $S_{[2]}$) for the different compositions:

Number of successes	p^2q			p^3q		p^2q^2		p^4q		p^3q^2	
	f	$S_{[1]}$	$S_{[2]}$	f	$S_{[2]}$	f	$S_{[2]}$	f	$S_{[2]}$	f	$S_{[2]}$
0	1	0	0	1	0	1	0	1	0	1	0
1	1	1	0	1	0	0	0	1	0	0	0
2	1	2	2	1	2	4	8	1	2	3	6
3	..	..	..	1	6	0	0	2	12	3	18
4	..	..	..	..	..	1	12	..	..	3	36
Total ..	3	3	2	4	8	6	20	5	14	10	60

This table gives the values of $\mu'_{[1]}$ and $\mu'_{[2]}$ for $m = 3, 4, 5$ for varying compositions. Using them in (I) and (II) the k's can be evaluated step by step.

Thus for example for $m = 3$, having the composition (21)

$$\mu_{[1]}' = \frac{3}{3} = k_1' \{p^2q\} \text{ or } 1 = k_1' \frac{2.1.1}{3.2.1},$$

as the probability of obtaining p^2q from three such points is $\frac{2.1.1}{3.2.1}$ or $k_1' = 3 = k_1''$

$$\frac{\mu'_{[2]}}{2!} = \frac{2}{3.2!} = k_3' \{p^2q\} = k_3' \frac{2.1.1}{3.2.1} \text{ or } k_3' = 1 = k_3''.$$

Similarly, for $m = 4$ having the composition (31)

$$\frac{\mu'_{[2]}}{2!} = \frac{2}{2!} = 2.k_3' \{p^2q\} + 1.k_4' \{p^3q\}$$

$$\text{or } 1 = 2.1.\frac{3.2.1}{4.3.2} + k_4' . \frac{3.2.1.1}{4.3.2.1} \text{ or } k_4' = 2 = k_4''.$$

Proceeding in a similar manner for the other k's we get

$$k_4''' = 6,\ k_5' = 0 = k_5'' \text{ and } k_5''' = 5 = k_5''''.$$

Substituting these in (I) and (II) and simplifying then for $\mu_{[1]}'$ and μ_2 we find that

$$\mu_{[1]}' = 3\,(m - 2)\,pq \text{ and } \mu_2 = (9m - 22)\,pq - (31m - 92)\,p^2q^2.$$

I am greatly indebted to Dr. P. V. Krishna Iyer for his valuable guidance and keen interest in the work.

References

1. Krishna Iyer, P. V. .. "The theory of probability distributions of points on a line," *J. Ind. Soc. Agr. Stat.*, 1948, **1**.
2. ——————— .. "Calculation of factorial moments of certain probability distributions," *Nature*, 1949, **164**.
3. ——————— .. "Further contributions to the theory of probability distributions of points on a line—I," *J. Ind. Soc. Agr. Stat.*, 1950, **2**.
4. ——————— .. "Further contributions to the theory of probability distributions of points on a line—II," *Ibid.*, 1951, **2**.
5. Uspensky, J. V. .. *Introduction to Mathematical Probability*, 1937.