ON THE LIMIT POINTS OF RELATIVE FREQUENCIES

By D. BASU

Statistical Laboratory, Calcutta

1. Summary

In the course of examining the alleged difficulty (J. Singh, 1946, Sankhyā, 7, 257-262) with von Mises definition of probability it was found that the alternative definition proposed by Singh does not make any difference in the situation as the existence of a 'quasi-limit', as defined by Singh, implies the existence of von Mises limit. A few interesting properties of the sequence of frequency ratios have also been considered.

2. Introduction

Let us consider an experiment the outcomes of which are classified into two categories, success and failure. The problem is to define the probability of obtaining a success from the experiment. Now if we repeat the experiment under identical set up a very large number of times we may get a sequence of 0's and 1's like, say, 01100011010011101...where 0 stands for a failure and 1 stands for a success.

Let $a(n)$ be the number of 1's (*i.e.* success) in the first n terms of the sequence. It is a common experience that the frequency ratio $a(n)/n$ becomes very stable for large n. Thus it is very natural to postulate that the frequency ratio $a(n)/n$ tends to a limit p as $n \to \infty$ and then to define the probability of success by the constant p. This is the well known approach of von Mises to the controversial problem of defining probability.

Singh tries to extend the scope of von Mises' theory by allowing for the existence of more than one limit points in the sequence $\{a(n)/n\}$ and postulating the existence of a limit point (called the 'quasi limit') round which almost all members of the sequence $\{a(n)/n\}$ clusters. The main result proved in this note is that the existence of 'quasi limit' implies the existence of the limit of the sequence $\{a(n)/n\}$. It is not the purpose here to enter into a controversy over the problem of defining probability. But it may be noted in passing that the objection raised by Singh against von Mises' theory is wrong and is based on a confusion between convergence in every particular collective with uniform convergence over all the possible collectives that can be generated by the random experiment.

3. Definitions

Singh defines a quasi limit q as follows:

Definition 1: Given two positive numbers ϵ, η as small as we please, then two numbers n_0 and m_0 can be found such that, of the m terms $a(n+i)/n+i$ $(i=1,2,\ldots m)$, π terms satisfy the inequality

$$\left| \frac{a(n+i)}{n+i} - q \right| < \epsilon \qquad \ldots \quad (3.1)$$

where $\pi/m > 1-\eta$ whenever $n \geqslant n_0$, $m \geqslant m_0$, the value of m_0 depending on ϵ, η and n.

It is apparent that the above definition somewhat relaxes the requirements for the existence of lim $a(n)/n$ for in that case given any $\epsilon>0$ we can always find an $n_0 = n_0(\epsilon)$ such that

$$|a(n)/n-q| < \epsilon \text{ for all } n \geqslant n_0 \quad \dots \quad (3.2)$$

We shall however prove that the quasi limit can exist when and only when lim $a(n)/n$ exists. But before that we give a definition of zero density of a sub sequence and an alternative simpler definition of a quasi limit.

Definition 2: The sub-sequence $\{\beta_n\}$ of the sequence $\{\alpha_n\}$ is said to have zero density in $\{\alpha_n\}$ if $g(n)/n \to 0$ as $n \to \infty$ where $g(n)$ is the number of β's in the first n terms of $\{\alpha_n\}$ (here $\{\beta_n\}$ is supposed to be obtained by deleting some terms of $\{\alpha_n\}$ and not by any alterations in the order of $\{\alpha_n\}$).

Definition 3: The sequence $\{a(n)/n\}$ is said to have a quasi limit q if for every $\epsilon>0$, no matter how small, the sub-sequence of $\{a(n)/n\}$ that satisfy the inequality

$$| a(n)/n - q | > \epsilon \quad \dots \quad (3.3)$$

is of density zero.

If lim $a(n)/n = q$ then the inequality (3.3) is satisfied for only a finite number of n's and the density of a finite sub-sequence is obviously zero. That definitions 1 and 3 are identical is seen as follows. If quasi limit q exists then from definition 1 it follows that given ϵ, η we can always find an n_0 such that in the sequence $\{a(n_0+i)/n_0+i)(i=1,2,\dots\text{ad inf})$ the proportion of those that satisfy (3.3) is ultimately $<\eta$ and remains $<\eta$. It then easily follows that in the entire sequence $\{a(n)/n\}(n=1,2,\dots\text{ad inf})$ the proportion of those that satisfy (3.3) ultimately remains $<\eta$ and since η is arbitrary it follows that the density of the sub-sequence is zero.

4. The main theorem

We now prove the main theorem namely

Theorem 1: *A quasi limit of the sequence $\{a(n)/n\}$ can exist if and only if lim $a(n)/n$ exists and in that case the two limits are identical.*

We first note that the difference between two consecutive terms $a(n)/n$ and $a(n+1)/n+1$ is $<1/n$. For

$$\left| \frac{a(n)}{n} - \frac{a(n+1)}{n+1} \right| = \frac{a(n)}{n(n+1)} \quad \text{or} \quad \frac{n-a(n)}{n(n+1)} \quad \dots \quad (4.1)$$

according as $a(n+1) = a(n)$ or $a(n)+1$.

Now we show that if q is a quasi limit of $\{a(n)/n\}$ then q is the only limit point of the sequence. Let us suppose that another limit point p exists such that $p > q$. Let $\epsilon > 0$ be an arbitrarily small quantity such that $\epsilon < \frac{1}{2}(p-q)$ and let $\alpha = p-q-2\epsilon > 0$.

Now since p is a limit point of $\{a(n)/n\}$ it follows that there exist an infinity of n's say $m_i(i = 1, 2, \ldots$ ad inf) such that

$$a(m_i)/m_i > p-\epsilon \quad (i = 1, 2, \ldots \text{ ad inf.}) \qquad \ldots \quad (4.2)$$

Since the difference between any two consecutive terms of $\{a(n)/n\}$ $(n \geqslant m_i)$ is $< \frac{1}{m_i}$ it follows that just after m_i at least $[\alpha m_i]$ terms satisfy the inequality

$$a(n)/n > q+\epsilon \qquad \ldots \quad (4.3)$$

where $[\alpha m_i]$ stands for the greatest integer not exceeding αm_i.

Thus in the first $m_i+[\alpha m_i]$ terms of the sequence $\{a(n)/n\}$ there are at least $[\alpha m_i]$ terms for which inequality (3.3) is satisfied. If $g(n)$ stands for the number of such terms in the first n terms of $\{a(n)/n\}$ then we have

$$g(m_i+[\alpha m_i]) \geqslant [\alpha m_i] \quad \text{for all } i$$

or

$$\frac{g(m_i+[\alpha m_i])}{m_i+[\alpha m_i]} \geqslant \frac{[\alpha m_i]}{m_i+[\alpha m_i]} \to \frac{\alpha}{1+\alpha} > 0 \quad \text{as } i \to \infty.$$

Thus for the sub-sequence $n = m_i+[\alpha m_i]$ $(i = 1,2,\ldots\text{ad inf.})$, $g(n)/n$ does not tend to zero and thus the definition of a qasi limit is violated. The case where $p<q$ can be similarly dealt with. Hence if a quasi limit q exists then there cannot exist any other limit point of the sequence $\{a(n)/n\}$ and thus theorem 1 is proved. It is to be noted that theorem 1 is true for the particular sequence $\{a(n)/n\}$ and not for a general sequence.

5. Theorem and example of incidental interest

The following theorem and example are of incidental interest.

Theorem 2: *If* $\lambda = \underline{\lim} \frac{a(n)}{n} < \overline{\lim} \frac{a(n)}{n} = \mu$ *then every point in the interval* (λ, μ) *is a limit point of the sequence* $\{a(n)/n\}$.

Since the sequence $\{a(n)/n\}$ is bounded both the limit inferior and the limit superior exist.

Let $\epsilon>0$ be an arbitrarily small number. Since λ and μ are both limit points of $\{a(n)/n\}$ it follows that both the inequalities

$$a(n)/n < \lambda+\epsilon \qquad \ldots \quad (5.1)$$

and

$$a(n)/n > \mu-\epsilon \qquad \ldots \quad (5.2)$$

are satisfied for an infinity of n's.

Hence we can always find two integers m_1 and m_2 such that $m_1<m_2$ and $1/m_1<\epsilon$ so that (5.1) and (5.2) are satisfied for $n = m_1$ and m_2 respectively.

Since $1/m_1<\epsilon$ it follows that the ratios $a(n)/n$ for $m_1 \leqslant n \leqslant m_2$ are ϵ-dense in the iterval (λ, μ) and since ϵ is arbitrary it follows that the sequence $\{a(n)/n\}$ is everywhere dense in the interval (λ, μ). Thus theorem 2 is proved.

That there can exist a sequence of 0's and 1's where the corresponding seqence of frequency ratios $\{a(n)/n\}$ are everywhere dense in the interval (0,1) can be easily seen from the following simple example.

Example: Let there be 1^2 one then 2^4 zeroes, then 3^6 ones then 4^8 zeroes and so on, in general $(2k+1)^{4k+2}$ ones follow $(2k)^{4k}$ zeroes and then is followed by $(2k+2)^{4k+4}$ zeroes. Now if we stop just at the end of the kth run of zeroes then clearly

$$\frac{a(n)}{n} = \frac{1^2+3^6+5^{10}+\ldots+(2k-1)^{4k-2}}{1^2+2^4+3^6+\ldots+(2k)^{4k}}$$

$$< \frac{k(2k-1)^{4k-2}}{(2k)^{4k}} \to 0 \text{ as } k \to \infty.$$

Again if we stop just at the end of the $(k+1)^{th}$ run of ones we have

$$\frac{a(n)}{n} = \frac{1^2+3^6+\ldots+(2k+1)^{4k+2}}{1^2+2^4+\ldots+(2k)^{4k}+(2k+1)^{4k+2}}$$

$$> \frac{(2k+1)^{4k+2}}{2k.(2k)^{4k}+(2k+1)^{4k+2}} \to 1 \text{ as } k \to \infty.$$

Thus $\underline{\lim}\, \frac{a(n)}{n} = 0$ and $\overline{\lim}\, \frac{a(n)}{n} = 1$ and therefore from theorem 2 every point in the interval (0, 1) is a limit point of $\{a(n)/n\}$.

My thanks are due to Dr. C. R. Rao and Mr. D. B. Lahiri for drawing my attention to the subject of this investigation.

Paper received : June, 1951.