The Time Distribution of So-Called Random Events

ARTHUR RUARK
University of North Carolina, Chapel Hill, North Carolina
(Received October 9, 1939)

It is shown that in a sequence of random events, the expected distribution of interval-sizes in a run of given duration differs from the Bateman distribution when the counting rate is different from the average one. The formula for "local" size-distribution to be expected for such a run is derived by the use of Bayes' theorem. It is pointed out that this theorem is useful in discussing the properties of any statistical system which is known to be in a condition differing from the idealized equilibrium state.

1. Description of a Fluctuation Problem of Very General Type

In analyzing the results of some counting experiments the writer encountered a statistical problem which appears to have an interest far transcending its immediate applications. It is in fact an instance of the following general problem:

A physical system governed by statistical laws is described by a number of variables a, b, $\cdots$, etc., both independent and dependent. If the system were in equilibrium they would take the values $\bar{A}$, $\bar{B}$, etc. In actuality the system fluctuates around the equilibrium state. In a particular experiment we find that a takes the deviant value A; or more accurately, we determine the probability αdA that a lies in the range A to $A+dA$. We inquire, what can be inferred as to the probabilities βdB, γdC, etc., that the variables b, c, etc. lay in ranges dB, dC, etc. at the time of the experiment?

Questions of this kind arising in the statistical mechanics of gases might be attacked by solving the Boltzmann differential equation for the velocity-distribution function, subject to the restriction $a=A$, but we want to point out that when the functions β, γ, etc. do not yield to direct attack, another method is available, namely, the use of Bayes' theorem of *a posteriori* probabilities. For example, if there are just two variables a and b to consider, this theorem states:

(Unconditional chance that a lies in dA)
$\times$(Chance that b lies in dB when we know that a lies in dA)
$=$(Unconditional chance that b lies in dB)
$\times$(Chance that a lies in dA when we know that b lies in dB).

The importance of such problems lies in the fact that their solution gives us a closer understanding of the conditions existing in the fluctuant states which are always encountered in nature. We proceed to the problem in radioactive fluctuations which led to these remarks. It will be solved by Bayes' theorem, for attempts to proceed otherwise yielded no valuable results.

2. A Problem in the Distribution of Random Events

A distribution of events is said to be random in time if the chance that one of them occurs in time dt is a constant, fdt. Bateman's formula gives the chance W_n that n events will occur in time t, namely,

$$W_n=(ft)^n e^{-ft}/n! \tag{1}$$

Then the chance that an interval between two events will exceed t is

$$W_0=e^{-ft}. \tag{2}$$

In deriving (1), all values of n up to "infinity" are supposed to be possible, and the understanding is that the formula shall be applied to a hypothetical situation,—that in which similar counting experiments, each of duration t, have been accumulated in "unlimited numbers." Similarly, (2) is a mathematical idealization referring to a single run of "unlimited duration." Here we shall discuss the physical situation encountered in studying long runs of events, whose distribution may be expected to conform roughly to the predictions of Eq. (2). Suppose we arrange an apparatus which does two things. It records the number of particles, n, given out by a constant radioactive source in a long time D, let us say 5000 seconds. Also, as each particle arrives, the instrument decides automatically whether the interval which has elapsed since the arrival of the preceding particle is greater than or less than a chosen value T (let us say one second). If the interval is greater than T, it is counted on a recording dial, so that at the end of the run we also know P, the number of intervals between events which exceed the value T. At this stage the best estimate we can make as to the average counting rate f is n/D,[1] but let us continue to record the total number of events, until f is known with very high accuracy. We shall assume for the present that f is exactly known, discussing later the consequences of the fact that f is really subject to statistical error. Now the problem we want to solve is this:

The average counting rate is f and a certain run has yielded n counts in time D. n is *not* equal to the expected value fD. What is the distribution function $G(t)$ replacing (2) during the interval D?[2]

3. The Solution

Figure 1 represents a chronograph sheet showing the events which arrived in the run of duration D. $G(T)$ is the chance that an interval between adjacent events, picked at random from the run, will be longer than T (and less than D, of course), when it is known that exactly n events arrived during the run. Bayes' theorem then says:

(Unconditional chance there are n events in the run)$\times G(T)=$(Unconditional chance an interval picked at random from the run will be longer than T and less than D)$\times$(Chance that n events occur in the run, when it is known that an interval picked at random from the run is longer than T and less than D).

The words "picked at random" are defined to mean that in the selection each interval has an

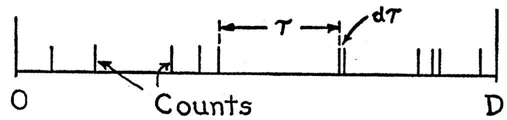

Fig. 1.

equal chance of being picked. For example, they might be assigned labels, and the labels could be thrown into an urn, after which one label would be drawn. In evaluating the unconditional chance in the right-hand member we must remember that all the intervals considered are less than D; for this subclass of intervals it is easy to show[3] that the probability of a length between T and D is $(e^{-fT}-e^{-fD})/(1-e^{-fD})$. Calling the last factor in Bayes' theorem φ, we have

$$G(T)=\frac{e^{-fT}-e^{-fD}}{1-e^{-fD}}\cdot\frac{\varphi}{(fD)^n e^{-fD}/n!}. \qquad (3)$$

In getting φ, we must remember the exact conditions: an interval I has already been picked out at random and has turned out to have a length between T and D, as shown in Fig. 1. An interval satisfying these conditions belongs to a subclass different from the subclass considered above, and the chance that its length lies in a range $d\tau$ is

$$e^{-f\tau}fd\tau/(e^{-fT}-e^{-fD}). \qquad (4)$$

A slight approximation will now be made. When fD is large we can afford to neglect the cases in which I is the first or the last interval in the run. These end intervals are each bounded by a single count, for the chance is zero that a count occurs exactly at time zero or at time D. Leaving these two intervals out of consideration, the very fact that we have picked an interior interval assures us that the run contains at least two counts, namely, the ones which initiate and terminate I. We are interested, then, in the chance that the two stretches outside I, which have a total length of $D-\tau$, shall contain just $(n-2)$ counts in all. This chance is just the same as though the intervals were contiguous, namely

$$[f(D-\tau)]^{n-2}e^{-f(D-\tau)}/(n-2)! \qquad (5)$$

Taking the product of (4) and (5), and integrat-

[1] Strictly speaking, it is $(n+1)/D$, as Meixner has shown. See reference 5.

[2] The law replacing Eq. (1) can be worked out from that replacing Eq. (2). See Ruark and Devol, Phys. Rev. **49**, 355 (1936); especially p. 356.

At first sight it seems that the problem we have set up differs from the general ones discussed above. The only difference is that we are seeking the integrated distribution $G(t)$ while $(dG/dt)dt$ is the chance that an interval lies between t and $t+dt$; this corresponds to βdB in the discussion above.

[3] A. Ruark and L. Devol, Phys. Rev. **49**, 355 (1936); see Section A-1.

ing over τ, we get

$$\varphi \cong \frac{[f(D-T)]^{n-1}}{(e^{f(D-T)}-1)(n-1)!}; \quad n=2, 3, \cdots; \tag{6}$$

$$G \cong \frac{1}{1-e^{-fD}} \cdot \frac{n}{fD} \cdot \left(1-\frac{T}{D}\right)^{n-1}; \quad n=2, 3, \cdots. \tag{7}$$

Here the approximation sign is used because of the simplification introduced above.[4] Neglecting the first fraction we rewrite (7) in the form

$$G \cong \frac{n}{fD}\left(1-\frac{T}{D}\right)^{n-fD-1} \cdot \left(1-\frac{fT}{fD}\right)^{fD}. \tag{7'}$$

Here the last factor approximates e^{-fT} very closely, so the first two factors show the effect of a known fluctuation in counting rate on the probability than an interval has a length greater than T. The first factor increases with n but the second decreases, and the rapidity of its fall depends on T. Either factor may predominate, depending on the value of T. It is easy to explain this physically. Suppose we have a set of counts, fD in number, which obey Eq. (2) very closely. We insert an extra count at random. By choosing fT large enough, we can make it certain that the total duration of the intervals shorter than T is large compared to the total duration of those longer than T. In such a case, there is a large probability that the extra count will fall into an interval shorter than T, splitting it into two still shorter ones and thereby decreasing the fraction of the intervals which are longer than T. The opposite conclusion is reached when fT is chosen very small. Detailed analysis shows that when fT is unity we are at the border between the two cases and G is insensitive to changes of n.

We next consider the influence of the fact that f itself is not accurately known. Suppose f has been determined by taking ν counts over a time t which is much longer than D. Meixner[5] showed that the chance f lies in df is

$$((ft)^{\nu}/\nu!)e^{-ft}tdf. \tag{8}$$

With this weight-factor the average of (7′) over all values of f[6] is found to be

$$\bar{G} = \frac{t}{\nu}\frac{n}{D}\left(1-\frac{T}{D}\right)^{n-1}. \tag{9}$$

The expected number of intervals, $\bar{P}$, having length greater than T is therefore $n\bar{G}$. One wishes also to know the standard deviation of the numbers P_1, P_2, etc. encountered in a series of runs. A direct attempt to evaluate this standard deviation leads to a result so complicated that it is useless, but when $(n-fD)$ is a small fraction of fD we can get a useful approximation.[7] Dropping the restriction that the counts must occur in a time D, we ask for the probability that in a run of n counts, P of the intervals are longer than T.

This is merely

$$W(n, P) = C_p^n e^{-fTP}(1-e^{-fT})^{n-P}. \tag{10}$$

On this simple basis, considering a large group of runs, the fractional standard deviation of P is $[(1-e^{-fT})/ne^{-fT}]^{\frac{1}{2}}$, and in applications this may be replaced by $\{[1-(P/n)]/P\}^{\frac{1}{2}}$.

To summarize: if we deal with a limited portion of a random distribution and find that the local counting rate is higher or lower than the average, the length-distribution law of the intervals differs from Eq. (2), and the expected local form of this distribution is given by Eq. (7), which is fundamental in the sense that it determines the expected local values of all properties of the sequence of events.

[4] Therefore Eq. (7) really gives the chance that an *interior interval* is greater than T when there are n counts in the run. The accurate formula for G can be obtained, but the situation is complex, so we prefer to work with (7).

[5] J. Meixner, Ann. d. Physik **30**, 665 (1937).

[6] Equation (7) can be averaged over f with the aid of a formula on p. 243 of Whittaker and Watson, *Modern Analysis*, fourth edition. The result is a series which converges uniformly for all physically possible values of its variable, and in fact very rapidly when ν is large.

[7] This has already been used in the doctoral dissertation of Dr. Mary W. Hodge, not yet published.