

STATISTICS OF A SET OF CLOSED INTERVALS

BY F. H. MURRAY

1. *Introduction.* In several types of engineering problems it is necessary to consider a group of infinite sequences of time intervals

$$t_{2i+1}^{(k)} \leq t \leq t_{2i+2}^{(k)}, \quad (k = 1, 2, \dots, m),$$

concerning which little is known except the average number of intervals per unit time, and the frequency distribution of the lengths, for each sequence. Examples are telephone calls being handled in an exchange, and many problems associated with the blowing of fuses.* It is proposed here to develop certain general formulas for a single sequence, and applications to a group of sequences in which the intervals of two distinct sequences are mutually independent.

2. *Non-Overlapping Intervals.* Consider any set of non-overlapping intervals

$$t' \leq t \leq t'',$$

and let $2Tn_T(l)$ be the number of intervals of the set each of length \bar{l} greater than or equal to l , whose end points are interior to the interval $-T < t < T$. If from each of these intervals an interval of length l is subtracted on the right, ($t_1'' = t_1'' - l$), the sum of the lengths of the parts remaining is a function of l , $2T\phi_T(l)$:

$$(1) \quad \phi_T(l) = \frac{1}{2T} \sum_{-T < t', t'' < T} [\bar{l}(\geq l) - l] = - \int_l^\infty (s - l) dn_T(s).$$

The functions $n_T(l)$, $\phi_T(l)$ are non-increasing; and since

$$2Tn_T(l) \leq \sum_{-T < t', t'' < T} (\bar{l} \geq l) \leq 2T,$$

it results that $n_T(l) \leq 1/l$.

Consider any sequence $T_1 < T_2 < \dots < T_n < \dots$, $T_n \rightarrow \infty$. Since $n_{T_k}(l)$ is bounded for each $l > 0$, if l has a rational value a sub-sequence ($T_{k'}$) exists such that

* See T. C. Fry, *Probability and its Engineering Uses*, Chapter 10.

$$\lim_{T_{k'} \rightarrow \infty} n_{T_{k'}}(l) = n(l).$$

Then on the set of positive rational numbers not including zero the Cantor diagonal method defines a sequence $(T_{k'})$ and a function $n(l)$, l rational, for which

$$\lim_{T_{k'} \rightarrow \infty} n_{T_{k'}}(l) = n'(l).$$

Since $n_T(l)$ is non-increasing, the sequence $(T_{k'})$ also defines a limit function at the points of continuity of $n(l)$.* The points of discontinuity of $n(l)$ form a denumerable set; for since $n_T(l) \leq 1/l$, the points of discontinuity for which $l \geq 2^{-m}$ form a denumerable set; and the set of discontinuities in each interval $2^{-m-1} \leq l < 2^{-m}$ is a denumerable set. The Cantor diagonal method can therefore be employed to define a subsequence $n_{T_k}(l)$ which converges at the points of discontinuity also; and

$$\lim_{T_{k''} \rightarrow \infty} n_{T_{k''}}(l) = n(l), \quad l > 0,$$

$$ln(l) \leq 1.$$

Since $n_T(l)$ is non-increasing and bounded for $l \geq l_1 > 0$, the theorem of Helly can be applied to (1), if $n_T(l) = 0, l > L$:

$$(2) \quad \lim_{T_{k''} \rightarrow \infty} p_{T_{k''}}(l) = p(l) = - \int_l^L (s - l) dn(s).$$

Integrating by parts, we find

$$(3) \quad p(l) = \int_l^L n(s) ds,$$

whence we see that $p(l)$ is continuous.

If no constant L exists for which $n_T(l) = 0, l > L$, an integration by parts of (1) gives

$$p_{T_{k''}}(l) = \int_l^\infty n_{T_{k''}}(s) ds;$$

and if $l < l'$,

$$-\Delta_l^{l'} p_{T_{k''}} = p_{T_{k''}}(l) - p_{T_{k''}}(l') = \int_l^{l'} n_{T_{k''}}(s) ds.$$

* For a complete discussion of a similar question, see A. Wintner, *Spektral Theorie der Unendliche Matrizen*, pp. 80-83; Leipzig, 1929.

The theorem of Lebesgue on the passage to the limit under the integral sign gives

$$(4) \quad -\Delta_l' p = \int_l^{l'} n(s) ds.$$

Since for every T ,

$$0 \leq p_T(l) - p_T(l') \leq 1,$$

we have also $0 \leq -\Delta_l' p \leq 1$, and as this is a non-decreasing function of l' the integral

$$(5) \quad \int_l^{\infty} n(s) ds$$

converges. This may be taken as the definition of $p(l)$. A sufficient condition for convergence to the limit is the assumption that for any $\epsilon > 0$, an L exists such that

$$(a) \quad p_T(l) < \epsilon, \quad l > L.$$

A related function is $f(l)$, defined by

$$(6) \quad f_T(l) = \frac{1}{2T} \sum_{-T < l', l'' < T} \bar{l}(\geq l) = - \int_l^{\infty} s dn_T(s);$$

from (1),

$$p_T(l) = f_T(l) - ln_T(l).$$

Hence if $p_{T_k''}$ and $n_{T_k''}$ approach limits, $f_{T_k''}$ approaches a limit function $f(l)$, and

$$(7) \quad p(l) = f(l) - ln(l).$$

If $p(l)$ is defined as above when hypothesis (a) is not satisfied, $f(l)$ may also be defined by this equation. In either case, $n(l+0)$ and $n(l-0)$ exist, and from (3)

$$n(l+0) = - \frac{d^+ p(l)}{dl}, \quad n(l-0) = - \frac{d^- p(l)}{dl}, \quad l > 0.$$

Since $p(l) \leq 1$, the integral (5) converges if the lower limit is zero; $f(l)$ is also bounded and non-increasing; hence $f(0)$ exists. If $n(0)$ exists, $p(0) = f(0)$; in any case

$$p(0) = f(0) - \lim_{l \rightarrow 0} ln(l).$$

3. *Several Sets of Intervals.* In what follows it will be assumed that the set of intervals is such that the functions $f(l)$, $n(l)$, $p(l)$ are uniquely defined as T becomes infinite, while $n(0) = n$ is finite. In this case n may be interpreted as the average or expected number of intervals per unit time; the function $g(l) = n(l)/n$ is the cumulative frequency function for the lengths of the intervals. The function $f(l)$ is the probability that a point selected at random falls on an interval of length at least l , and $p(l)$ is the probability that a random point t^0 , with all points of the interval $(t^0, t^0 + l)$, lies on some interval of the sequence. The number n and the function $g(l)$ determine the remaining functions, and in many cases are the quantities most easily measured.

Suppose two sets present, having end-points (α', β') , (α'', β'') respectively, and assume that the intervals of each set are independent of the other. The set of intervals common to both sets may be divided into two types; in the first, the initial point α' of the first set lies on an interval (α'', β'') , and in the second α'' lies on an interval (α', β') . The average number of intervals of length at least Δ , of type (1), is equal to $n'(\Delta)p''(\Delta)$, or the number $n'(\Delta)$ of intervals (α', β') times the probability that α' lies on an interval (α'', β'') together with the interval $(\alpha', \alpha' + \Delta)$. The number of intervals of type (2) is $n''(\Delta)p'(\Delta)$; hence

$$N_{12}(\Delta) = n'(\Delta)p''(\Delta) + n''(\Delta)p'(\Delta)$$

is the expected number per unit time. Also, from the general formulas,

$$\begin{aligned} P_{12}(\Delta) &= \int_{\Delta}^{\infty} N_{12}(s) ds \\ &= - \int_{\Delta}^{\infty} [p''(s) dp'(s) + p'(s) dp''(s)] \\ &= - \left[p'(s)p''(s) \right]_{\Delta}^{\infty} = p'(\Delta)p''(\Delta), \end{aligned}$$

if the variation of each of the functions (p', p'') over the set of discontinuities of the other is zero.*

* See Hobson, *Functions of a Real Variable*, vol. 1, p. 542, 1927. In many applications this condition can be satisfied by redefining, if necessary, the functions p', p'' in such a way that the original functions are obtained by a passage to the limit.

If there are m sets of intervals, the function $P_{12\dots m}(\Delta)$ is proved by induction to be equal to $p'(\Delta)p''(\Delta) \cdots p^{(m)}(\Delta)$. The number of such intervals per unit time, of length $\geq \Delta$, is found by differentiation, and the function $F_{12\dots m}(\Delta)$ is also determined by the general formulas. The function $N_{12\dots m}(\Delta)$ can have discontinuities at the values $\Delta = l_i$ which are discontinuities for any of the original sequences, and only at these points, since

$$\frac{d}{d\Delta}P_{12\dots m}(\Delta)$$

is uniquely determined everywhere else.

4. *Intervals of Constant Length.* If the intervals are of constant length h ,

$$\begin{aligned} n(l) &= n(0) = n, & l \leq h, \\ &= 0, & l > h, \\ p(l) &= n(h-l), & l \leq h. \end{aligned}$$

If m sequences are present, with the same n and h ,

$$\begin{aligned} P_m(l) &= p(l)^m = n^m(h-l)^m, & l \leq h, \\ N_m(l) &= mn^m(h-l)^{m-1}. \end{aligned}$$

The function

$$\frac{N_m(l)}{N_m(0)} = \left(1 - \frac{l}{h}\right)^{m-1}$$

represents the cumulative frequency distribution for the common lengths, of which the average length is found to be h/m .

If the function

$$g(l) = \frac{n(l)}{n}$$

varies exponentially, we have

$$\begin{aligned} n(l) &= ne^{-l/h}, \\ p(l) &= \int_1^\infty n(s)ds = nh e^{-l/h}; \end{aligned}$$

and if m similar sequences are present, we have

$$P_m(l) = (nh)^m e^{-ml/h},$$

$$N_m(l) = \frac{m}{h} (nh)^m e^{-ml/h},$$

$$G_m = \frac{N_m(l)}{N_m(0)} = e^{-ml/h}.$$

For a single sequence, the average holding time is

$$- \int_0^{\infty} l d e^{-l/h} = h.$$

The average length of an interval common to m sequences is

$$- \int_0^{\infty} l d G_m = \frac{h}{m}.$$

In all cases, the average length of an interval common to m sets is

$$\begin{aligned} \bar{l} &= - \int_0^{\infty} l d G_m(l) = - \frac{1}{N_m(0)} \int_0^{\infty} l d N_m(l) \\ &= \frac{F_m(0)}{N_m(0)} = - \frac{P_m(0)}{P_m'(0)}. \end{aligned}$$

This becomes

$$\begin{aligned} \frac{1}{\bar{l}_m'} &= - \left. \frac{d}{dl} (p_1 p_2 \cdots p_m) \right|_{l=0} \\ &= - \sum_{k=1}^m \frac{p_k'(0)}{p_k(0)} = \sum_{k=1}^m \frac{1}{\bar{l}_k}. \end{aligned}$$

In many applications the constant $p_k(0) = p_k$ can be interpreted as a load due to an individual source; the equation $\bar{l} = -p(0)/p'(0)$ becomes $\bar{l} = p/n$, and in the form $p = n\bar{l}$ is self-evident.