A STATISTICAL PROBLEM CONNECTED WITH THE COUNTING OF RADIOACTIVE PARTICLES

By Sten Malmquist

Institute of Statistics, University of Upsala, Sweden

1. Introduction. Our problem refers to random events forming a sequence in time or in space, *e.g.* particles emitted by a radioactive matter. By omitting certain elements of the given sequence, say f, we form another sequence, say g. The rule of omission involves an arbitrarily prescribed constant u. The rule to be followed in forming g is:

Case I: Let a be an element in f and g. The next element to be included in g is then the first element in f which follows a after a distance greater than u.

Case II: Let a be an element in f and g. The next element to be included in g is then the first element in f which follows a at a distance greater than u from the preceding element in f, whether this belongs to g or not.

When the events are represented by impulses emitted by a radioactive matter and feeding a recorder with a constant resolving time u, the new sequence consists of the counted impulses. The two cases correspond to the reaction of different types of recorders. The distinction between the two transformations has caused some confusion. It has, however, been clearly pointed out by Ruark and Brammer [5].

v. Bortkiewicz [2] seems to be the first who has considered problems related to the transformed sequence. Starting from investigations by Rutherford, Geiger, and others, concerning the number of recorded α-particles during a certain interval of time, say T, he observed that the distribution of this number was similar to that of Poisson but with a slightly smaller dispersion. This fact he supposed to be caused by a constant resolving time u of the recorder. By means of certain assumptions he tried to calculate the effect on the mean and the dispersion by the transformation in Case I, supposing the cumulative distribution function $F(t)$ for the distance between two consecutive elements in the sequence f is given by

$$F(t) = 1 - e^{-at},$$

where here and in what follows, t denotes a non-negative variable.

Considering Case II with $F(t)$ as above, Levert and Scheen [4] have recently worked out an expression for the distribution of the number of elements during T in the sequence g.

Gnedenko [3] has considered the distribution of the number of lost elements in Case I with particular regard to the initial state of rest.

Alaoglu and Smith [1] considered problems referring to successive transformations of a sequence. When, for example, a sequence of particles enters a tube-counter and amplifier, together acting with a resolving time u_1, and the impulses then are feeding a recorder with resolving time $u_2 > u_1$, the sequence of recorded impulses will be the result of two successive transformations. If we have a scaling circuit between the counter and the recorder, we have to make a transformation of another type between the two transformations in Case I and Case II.

The present paper deals with the transformed sequence in Case I. The distribution function $F(t)$ is supposed to be arbitrary. An advantage of this generalization is that the formulas derived could be used in treating problems referring to successive transformations.

The author wishes to express his sincere gratitude to Professor Herman Wold for stimulating discussions and valuable advice.

2. Derivation of distributions for case I. Suppose that the sequence f has $F(t)$ for distribution function for the distance between two consecutive elements. $F(t)$ is supposed to be independent of absolute time (space), and of the preceding distance between two elements. When not stated otherwise, we further suppose $F(0) = 0$.

Now let $G(t)$ be the distribution function for the distance between two consecutive elements in the transformed sequence g. Evidently $G(t)$ also is independent of absolute time and of the preceding distance between two elements.

We shall consider certain distribution functions connected with $F(t)$. These functions will then be used in solving problems concerning the sequence g.

Let $F_n(t)$ be the distribution function for the distance between the first and the last of $n + 1$ consecutive elements in the sequence f. Then $F_n(t)$ is given by the recursive system

$$(1) \qquad \begin{aligned} F_{m+n}(t) &= \int_0^t F_m(t - x)\, dF_n(x); \qquad (m, n \geqq 1) \\ F_1(t) &\equiv F(t). \end{aligned}$$

As is easily seen, we have

$$F_{m+n}(t) \leq F_m(t) \cdot F_n(t);$$

and, for $t = u$,

$$F_n(u) \to 0, \qquad \text{as } n \to \infty;$$

$$\sum_{n=1}^{\infty} F_n(u) < \infty, \qquad \text{provided that } F_1(0) < 1.$$

Alternatively, $F_n(t)$ could be deduced by the use of characteristic functions.

Still considering the sequence f, let $\Phi(t)$ be the distribution function for the distance d between an arbitrarily chosen point and the following element. Suppose that the arbitrary point is chosen so that the distance between the preceding and the following element is x. Under this condition we have, in usual symbols,

$$P(d > t) = \frac{x - t}{x}.$$

Hence,

$$\Phi(t) = 1 - \int_t^\infty \frac{x - t}{x} \, dH(x)$$

where $H(t)$ is the distribution function for the distance x.

To deduce $H(t)$ we suppose that the distribution $F(t)$ has a finite mean,

$$m = \int_0^\infty t \, dF(t).$$

By the definition of $H(t)$, we then have

$$H(x) = \frac{1}{m} \int_0^x t dF(t).$$

Thus

$$\Phi(t) = \frac{1}{m}\left[\int_0^t x \, dF(x) + t \int_t^\infty dF(x)\right]. \tag{2}$$

The corresponding frequency function $\varphi(t)$ is given by

$$\varphi(t) = \frac{1 - F(t)}{m}.$$

Consider $n + 2$ consecutive elements in f, say $a_0, a_1, \cdots, a_{n+1}$, where a_0 is an element in the transformed sequence g. The probability P_n that the next element in g following a_0 will be a_{n+1} is given by

$$P_n = F_n(u) - F_{n+1}(u), \qquad (n = 1, 2, \cdots),$$
$$P_0 = 1 - F(u).$$

Now let $P_n(t)$ be the probability that the distance between a_0 and a_{n+1} is smaller than or equal to t, when a_0 an a_{n+1} are two consecutive elements in the sequence g. Then

$$P_n(t) = \frac{1}{F_n(u) - F_{n+1}(u)} \int_0^u \left[F(t - x) - F(u - x)\right] dF_n(x),$$
$$(n = 1, 2 \cdots), \qquad P_0(t) = \frac{F(t) - F(u)}{1 - F(u)}.$$

Let $G^*(t)$ be defined by

$$G^*(t) = \sum_{n=0}^\infty P_n \cdot P_n(t) = F(t) - F(u)$$
$$+ \sum_{n=1}^\infty \int_0^u [F(t - x) - F(u - x)] \, dF_n(x); \qquad t > u.$$

When $G^*(t)$ is a distribution function, then $G^*(t)$ equals $G(t)$.

For $t_1 < t_2$ we obviously have $G^*(t_1) \leq G^*(t_2)$.

For $t = \infty$

$$G^*(\infty) = 1 - F(u) + \sum_{n=1}^{\infty} \int_0^u [1 - F(u - x)]\, dF_n(x)$$

$$= 1 - F(u) + \sum_1^{\infty} F_n(u) - \sum_1^{\infty} F_{n+1}(u) = 1.$$

Hence we take

$$G(t) = G^*(t); \qquad t > u, \tag{4}$$

$$G(t) = 0; \qquad t \leq u.$$

When the corresponding frequency functions $g(t)$ and $f(t)$ exist, we get

$$g(t) = f(t) + \sum_{n=1}^{\infty} \int_0^u f(t - x) f_n(x)\, dx; \qquad t > u. \tag{5}$$

Dealing with a sequence of elements we are often concerned with the number of occurrences during a certain time T.

Let the mean number of occurrences during T be $M(T)$. Supposing that the mean $m = \int_0^{\infty} t\, dF(t)$ is finite and that $F(0) < 1$, we have

$$M(T) = T/m. \tag{6}$$

We define

$$K_1(t) = \begin{cases} F(t) & \text{for } t \geq \epsilon \\ 0 & \text{for } t < \epsilon \end{cases}$$

$$K_2(t) = \begin{cases} F(t) & \text{for } t \geq \epsilon \\ F(\epsilon) & \text{for } t < \epsilon \end{cases}$$

and denote the corresponding means by $M_1(T)$ and $M_2(T)$. As is easily seen,

$$M_1(\epsilon) \leq M(\epsilon) \leq M_2(\epsilon).$$

Using (2),

$$M_1(\epsilon) = \frac{\epsilon F(\epsilon) + \epsilon[1 - F(\epsilon)]}{\int_0^{\infty} x\, dK_1(x)} = \frac{\epsilon}{\int_0^{\infty} x\, dK_1(x)};$$

$$M_2(\epsilon) = \frac{1}{\int_0^{\infty} x\, dK_2(x)} [1 \cdot \epsilon[1 - F(\epsilon)]^2 + \cdots + n \cdot \epsilon[1 - F(\epsilon)]^2 F(\epsilon)^{n-1} + \cdots]$$

$$= \frac{\epsilon}{\int_0^{\infty} x\, dK_2(x)}.$$

Making $N = T/\epsilon$ and summing, we obtain

$$M_1(T) = \frac{T}{\int_0^\infty x\, dK_1(x)} = \frac{T}{m - \int_0^\epsilon x\, dF(x) + \epsilon F(\epsilon)};$$

$$M_2(T) = \frac{T}{\int_0^\infty x\, dK_2(x)} = \frac{T}{m - \int_0^\epsilon x\, dF(x)}.$$

By choosing ϵ arbitrarily small, we get

$$M(T) \rightarrow T/m.$$

Let $P(n, T)$ be the probability that we get n elements in f during a time T. Suppose that the first of these elements, a_1, comes at $T_0 + x$, and the last, a_n, at $T_0 + x + y$.

We then have

$$(7) \qquad P(n, T) = \int_0^T \varphi(x)\, dx \int_0^{T-x} [1 - F(T - x - y)]\, dF_{n-1}(y).$$

In (4) and (7) we have equations for the transformation in Case I. Because of the general form of $F(t)$, the formulas also can be used when we are concerned with successive transformations. It can further be remarked that the transformation of a sequence of impulses by passing a scaling circuit is expressed by the system (1).

3. Results for a particular form for $F(t)$**.** The preceding formulas will now be used for a special distribution function $F(t)$. Suppose that the frequency function $f(t) = dF(t)/dt$ is equal to the frequency function of the distance between an arbitrary point and the following element.

From (3) we get

$$F'(t) = \frac{1 - F(t)}{m},$$

or, when $F(0) = 0$,

$$(8) \qquad F(t) = 1 - e^{-at};$$

$$(9) \qquad f(t) = ae^{-at}, \quad \text{where } 1/a = m = \int_0^\infty tf(t)\, dt.$$

By means of the theory of characteristic functions we have

$$(10) \qquad f_n(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} [\eta(x)]^n e^{-itx}\, dx; \qquad f_1(t) \equiv f(t);$$

where

$$(11) \qquad \eta(x) = a \int_0^\infty e^{-at} e^{itx}\, dt = \frac{a}{a - ix}.$$

Thus

$$f_n(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{a^n}{(a - ix)^n} e^{-itx} \, dx \tag{12}$$

For $n = 1$, we get

$$f_1(t) = ae^{-at} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{a}{a - ix} e^{-itx} \, dx \tag{13}$$

By differentiating (13) $n - 1$ times with respect to a we obtain

$$(-t)^{n-1} e^{-at} = \frac{1}{2\pi} (-1)^{n-1}(n - 1)! \int_{-\infty}^{+\infty} \frac{e^{-itx}}{(a - ix)^n} \, dx.$$

Hence, from (12),

$$f_n(t) = \frac{a^n}{(n - 1)!} t^{n-1} e^{-at}. \tag{14}$$

From (5) we obtain the frequency function for the transformed sequence g

$$\begin{aligned} g(t) &= ae^{-at} + \sum_{n=1}^{\infty} \int_0^u ae^{-at} \frac{a^n}{(n - 1)!} t^{n-1} \, dx = ae^{au} e^{-at}; \qquad t \geq u \\ G(t) &= 0; \; t < u. \end{aligned} \tag{15}$$

The mean m_g is given by

$$m_g = a \int_u^{\infty} te^{au} e^{-at} \, dt = \frac{1}{a} + u.$$

Remark: Suppose the constant u is allowed to vary independently of t and that the frequency function of u is $\gamma(u)$, we obtain

$$\begin{aligned} m_g &= \int_0^{\infty} t \, dt \int_0^t g(u, t)\gamma(u) \, du = \int_0^{\infty} \frac{1}{a} \gamma(u) \, du + \int_0^{\infty} u\gamma(u) \, du \\ &= \frac{1}{a} + m(u). \end{aligned} \tag{16}$$

Now let the sequence of elements, g, by means of (5) be transformed into a new sequence, h. When we are concerned with the counting of particles, emitted from a radioactive matter, let the sequence g consist of impulses from a counter-amplifier with resolving time u, feeding a recorder with resolving time u_1. Then the elements in h are the counted impulses, it being supposed that the tube-counter and the recorder reacts according to the assumptions.

We suppose $u_1 > u$. When $u_1 \leq u$, the sequences g and h are identical.

Let $g_n(t)$ denote the frequency function of the distance between the first and the last of $n + 1$ consecutive elements in g. We find, in the same way as used in obtaining (14),

$$g_n(t) = \frac{a^n}{(n - 1)!} e^{anu}(t - nu)^{n-1} e^{-at}; \qquad t \geq nu. \tag{17}$$

Let $h(t)$ be the frequency function for the distance between two consecutive elements in the sequence h. Let further N be the greatest integer smaller than or equal to u_1/u.

Using (4) and (5) we obtain

$$\text{(18)}\quad \begin{aligned} h_I(t) &= ae^{au}e^{-at}\sum_0^N \frac{a^n}{n!}(u_1 - nu)^n e^{anu}; && t \geq u_1 + u; \\ h_{II}(t) &= ae^{au}e^{-at}\sum_0^N \frac{a^n}{n!}[t - (n+1)u]^n e^{anu}, && (N+1)u \leq t \leq u_1 + u; \\ h_{III}(t) &= ae^{au}e^{-at}\sum_0^{N-1} \frac{a^n}{n!}[t - (n+1)u]^n e^{anu}, && u_1 \leq t \leq (N+1)u. \end{aligned}$$

The mean m_h is found to be

$$\text{(19)}\qquad m_h = \left[\frac{1}{a} + u\right]\left[1 + \sum_{n=1}^{N}\sum_{v=n}^{\infty}\frac{(u_1 - nu)^v a^v}{v!}e^{-a(u_1 - nu)}\right].$$

We also have

$$\int_{u_1+u}^{\infty} th_I(t)\,dt < m_h < \int_u^{\infty} th_I(t)\,dt$$

or

$$\left[\frac{1}{a} + u_1 + u\right]\left[\sum_0^N \frac{a^n}{n!}(u_1 - nu)^n e^{-a(u_1 - nu)}\right] < m_h < \left[\frac{1}{a} + u_1\right]e^{au}\left[\sum_0^N \frac{a^n}{n!}(u_1 - nu)^n e^{-a(u_1 - nu)}\right].$$

We now consider the number of occurrences during a time interval T. Using (6), (16), and (19) we immediately get the mean numbers of occurrences during T.

By (3), we get for the sequence g

$$\text{(20)}\qquad \varphi_g(t) = \begin{cases} \dfrac{a}{au+1}; & t \leq u \\ \dfrac{a}{au+1}e^{au}e^{-at}; & t \geq u. \end{cases}$$

Inserting (20), (15) and (14) in (7) and evaluating the integrals, we finally get

$$\text{(21)}\quad P_g(n, T) = \begin{cases} a_{n-1} - 2a_n + a_{n+1}; & n \leq \dfrac{T}{u} - 1 \\ a_{n-1} - 2a_n + (n+1) - \dfrac{aT}{au+1}; & \dfrac{T}{u} - 1 \leq n \leq \dfrac{T}{u} \\ a_{n-1} - 2\left[n - \dfrac{aT}{au+1}\right] + (n+1) - \dfrac{aT}{au+1}; & \\ & \dfrac{T}{u} \leq n \leq \dfrac{T}{u} + 1. \end{cases}$$

where

$$a_n = \frac{1}{au + 1} e^{-a(T-nu)} \sum_{v=0}^{n} \frac{(T - nu)^v a^v}{v!} (n - v), \qquad (n = 0, 1, \cdots), \tag{22}$$

$$a_{-1} = 0.$$

When $u = 0$, we obtain

$$a_n = e^{-aT} \sum_{v=0}^{n} \frac{T^v a^v}{v!} (n - v).$$

For the sequence f we then get the Poisson distribution

$$P_f(n, T) = \frac{(aT)^n}{n!} e^{-aT}. \tag{23}$$

The corresponding expression for the sequence h is much more complicated.

4. A statistical experiment. The following statistical experiment will serve as an illustration of the scheme dealt with in this paper—the transformation of a sequence and the resulting formulas, especially (21).

Groups of five figures, the last rounded up if necessary, have been extracted from tables of random sampling numbers (6). Let each group denote the first five digits for a decimal x, arbitrarily chosen between 0 and 1. The variable x is supposed to have the distribution function t for $0 \leq t < 1$. We now define a new variable, y, given by

$$y = -k \log (1 - x), \text{ [or } y = -k \log x]. \tag{24}$$

The variable y has the distribution function given by (8), *viz.*

$$F(t) = 1 - e^{-at}, \text{ where } \frac{1}{a} = m = k \log e.$$

Transforming each group, or number x, according to (24), we get a sample of consecutive distances between elements in the sequence f considered in the previous sections. Choosing a constant u, we can construct the corresponding sequence g. Beginning with a point, arbitrarily chosen on the first distance, we can finally count the number of elements in successive intervals of the same length.

Take $k = 1$, $u = 0.2$ and $T = 1.5$. We then have for the sequences f and g:

$$m_f = \frac{1}{a} = \log e = 0.4343; \qquad m_g = \frac{1}{a} + u = 0.6343;$$

$$\sigma_f = \frac{1}{a} = 0.4343; \qquad \sigma_g = \frac{1}{a} = 0.4343;$$

$$M_f(T) = \frac{T}{m_f} = 3.454. \qquad M_g(T) = \frac{T}{m_g} = 2.365.$$

The experiment yielded the following results:

For the sequence f:
Number of elements 801.
$\bar{m}_f = 0.450$.

For the sequence g:
Number of elements 555.
$\bar{m}_g = 0.648$.

In neither case is the deviation between the observed and theoretical means statistically significant. In fact we have:

$$\frac{(\bar{m}_f - m_f)\sqrt{800}}{\sigma_f} \sim 1.0; \qquad \frac{(\bar{m}_g - m_g)\sqrt{554}}{\sigma_g} \sim 0.8$$

which gives $P = 0.3$ and $P = 0.4$, respectively.

TABLE I

Nos. of intervals with n elements

n	Sequence f		Sequence g		
	Observed	Expected according to (23)	Observed	Expected according to (21)	Expected according to (23)
0	6	7.6	5	8.2	23.7
1	33	26.1	53	42.5	54.8
2	48	45.1	82	81.8	63.3
3	55	51.9	69	72.2	48.8
4	36	44.8	23	29.2	28.1
5	32	31.0	6	4.8	13.0
6	17	17.8	1	0.2	5.0
7–	12	14.7			2.4
Σ	239	239	239	238.9	239
Mean	3.331	3.454	2.310	2.36	2.31
χ^2		4.825		4.524	36.7
P		0.68		0.34	<0.001

The functions a_n in (22) can be calculated by means of Pearson's tables of the incomplete γ-function (7). In the notation of these tables we obtain

$$e^{-\lambda} \sum_{v=r}^{\infty} \frac{\lambda^v}{v!} = I\left(\frac{\lambda}{\sqrt{r-1}}; r-2\right) = I(p, q).$$

Hence

$$a_n = \frac{n}{au+1} e^{-\lambda} \frac{\lambda^n}{n!} + \frac{n-\lambda}{au+1} [1 - I(p, q)],$$

where

$$\lambda = a(T - nu); \qquad p = \frac{\lambda}{\sqrt{n-1}}; \qquad q = n - 2.$$

In the present case, however, we only need the numbers up to a_7. Accordingly, the a_n have been calculated directly.

The resulting theoretical and observed distributions for the number of elements during T for the sequences f and g will be found in Table I. For comparison, a Poisson distribution, with the same mean as observed for the sequence g_1 is given. The result of a χ^2 test is also shown in Table I. Judged by the χ^2 test the distributions (23) and (21) agree fairly well with the observed distributions. As was to be expected, the Poisson distribution cannot be used for the sequence g.

REFERENCES

[1] L. Alaoglu and N. M. Smith, "Statistical theory of a scaling circuit," *Phys. Rev.*, Vol. 53 (1938), pp. 832–836.

[2] L. v. Bortkiewicz, *Die radioaktive Strahlung als Gegenstand wahrscheinlichkeitstheoretischer Untersuchungen*, Berlin, 1913.

[3] B. V. Gnedenko, "On the theory of Geiger-Müller counters," (in Russian), *Jour. for Exp. and Theor. Phys.*, Vol. 11 (1941), pp. 101–106.

[4] C. Levert and W. L. Scheen, "Probability fluctuations of discharges in a Geiger-Müller counter produced by cosmic radiation," *Physica*, Vol. 10 (1943), pp. 225–238.

[5] A. E. Ruark and F. E. Brammer, "The efficiency of counters and counter circuits," *Phys. Rev.*, Vol. 52 (1937), pp. 322–324.

[6] M. G. Kendall and B. Babington Smith, *Tables of Random Sampling Numbers*, Tracts for Computers XXIV, Cambridge, 1939.

[7] Karl Pearson, *Tables of the Incomplete* Γ*-function*, Cambridge, 1922.