A USE OF COMPLEX PROBABILITIES IN THE THEORY OF STOCHASTIC PROCESSES

By D. R. COX

Received 17 September 1954

ABSTRACT. The exponential distribution is very important in the theory of stochastic processes with discrete states in continuous time. A. K. Erlang suggested a method of extending to other distributions methods that apply in the first instance only to exponential distributions. His idea is generalized to cover all distributions with rational Laplace transforms; this involves the formal use of complex transition probabilities. Properties of the method are considered.

1. The exponential probability density function, $\lambda e^{-\lambda t}$, plays an important role in the study of stochastic processes having discrete states in continuous time. For if the distribution of 'life-time' has the above form the probability of 'death' in a time interval δt is unconditionally $\lambda \delta t + o(\delta t)$. If all the life-times connected with a system have exponential distributions, the process is Markovian, and standard methods (4 yield a set of ordinary linear differential equations to determine the behaviour of the process and a set of simple linear equations to determine the equilibrium distribution, if one exists. Examples of life-times are the service-times of customers in a queue, the intervals between the arrivals of successive customers in a queue, the division-times of bacteria, and so on.

In his pioneer work on congestion in telephone systems Erlang (3) introduced a simple and ingenious method for covering some other distributions; he supposed that life has k stages, the times spent in each following independently the distribution $\lambda e^{-\lambda t}$. The total life-time is then distributed proportionally to χ^2 with $2k$ degrees of freedom and the process is Markovian, provided that the description of the state of the system includes an account of which stage of life has been reached. The division into stages is a mathematical device and need have no physical significance. Erlang's idea is particularly useful in fairly practical investigations because it gives a routine method of dealing with non-exponential distributions. Some examples of its use are (6), (7) and (2).

Jensen (5) has thoroughly discussed a generalization of Erlang's device, including the case when the λ's associated with the different stages are not equal. In the present paper we discuss further generalizations of the method.

2. The Laplace transform of a function $f(t)$ will be denoted by $f^*(s)$, i.e.

$$f^*(s) = \int_0^\infty e^{-st} f(t)\, dt. \tag{1}$$

If $f(t)$ is the probability-density function (p.d.f.) of a non-negative random variable we shall say that $f^*(s)$ is in $\mathscr{T}_+$; the integral for $f^*(s)$ will be convergent in at least the positive half-plane.

For the exponential distribution

$$f^*(s) = \lambda/(s+\lambda). \tag{2}$$

For the generalization of Erlang's device in which the λ's in the k stages are $\lambda_1, \ldots, \lambda_k$, the Laplace transform is

$$\prod_{i=1}^{k} \lambda_i/(s+\lambda_i), \tag{3}$$

because the p.d.f. of the total life-time is the convolution of the p.d.f.'s of the times spent in the separate stages.

Thus $f^*(s)$ is the reciprocal of a polynomial and has its poles on the negative real axis of the complex s-plane. The position of the poles determines the transition probabilities λ_i, and the number of poles, counted according to multiplicity, equals the number of stages.

As our first generalization, suppose that $f^*(s)$ is in $\mathscr{T}_+$ and is the reciprocal of a polynomial not necessarily with real zeros. Since $f^*(s)$ is real, complex poles must occur in conjugate pairs; also it is easily seen that the real parts of the poles must be negative, i.e. if $f^*(s)$ has the form (3), $\Re\,(\lambda_i) > 0$. Then, even though the transition densities λ_i are complex, we can, purely formally, still treat the process as Markovian and form the differential equations in the usual way. It can be shown that although in the solution the probabilities associated with the fictitious stages may be complex, the probabilities associated with real states of the system will be real.

The simplest case is $\lambda_1 = a$, $\lambda_2 = a+ib$, $\lambda_3 = a-ib$, with $a > 0$, when the Laplace transform is

$$f^*(s) = \frac{a(a^2+b^2)}{(a+s)\{(a+s)^2+b^2\}}, \tag{4}$$

so that

$$f(t) = ab^{-2}(a^2+b^2)\,e^{-at}(1-\cos bt). \tag{5}$$

If, for example, the second stage has been reached, the formal probability of passing on to the third in δt is $(a+ib)\,\delta t + o(\delta t)$.

3. The previous section dealt with Laplace transforms that are the reciprocals of polynomials. We now consider rational functions. Let $f^*(s)$ be in $\mathscr{T}_+$ and be the ratio of a polynomial of degree at most k to a polynomial of degree k. Then if the zeros of the denominator are at $-\lambda_i$ $(i = 1, \ldots, k)$, we can expand $f^*(s)$ in partial fractions in many ways, of which we still consider two, viz.

$$f^*(s) = w_0 + \sum_{i=1}^{k} \frac{w_i \lambda_i}{\lambda_i + s}, \tag{6}$$

where $\Sigma w_i = 1$ and the λ_i are all different, and, for the second expansion, with no restriction on λ_i,

$$f^*(s) = p_0 + \sum_{i=1}^{k} q_0 \cdots q_{i-1} p_i \prod_{l=1}^{i} (1+s/\lambda_l)^{-1}, \tag{7}$$

where p_i, q_i are constants, $p_i + q_i = 1$, $p_k = 1$.

Equation (6) is the simpler, but it is convenient to have the form (7) independent of restrictions on λ_i. Both have simple probability interpretations. For (6) there is a probability w_0 of immediate death and a probability w_i of beginning on the ith stage, life then consisting of a single stage. In equation (7) there is a probability p_0 of zero life and a probability q_0 of entering the first stage. In the first stage the probability of passing to the second stage in δt is $q_1 \lambda_1 \delta t + o(\delta t)$, and the probability of death in δt is $p_1 \lambda_1 \delta t + o(\delta t)$, and so on. Both these interpretations are purely formal in that we place no restriction on w_i, p_i, λ_i; our sole requirement is that $f^*(s)$ shall be in $\mathscr{T}_+$.

Two comments need to be made on (6) and (7). First, if the degree of the numerator of $f^*(s)$ were to exceed the degree of the denominator, $f^*(s)$ would be unbounded on the positive real axis and this is inadmissible. Secondly, we could consider a scheme (5) in which at the end of the ith stage, instead of a choice between death and the $(i+1)$th stage, death can occur or any stage of life entered or re-entered. We now prove that this does not lead to a class of distributions more general than (7).

Suppose that there are k stages of life and that in the ith stage the probability of death in δt is $d_i \lambda_i \delta t + o(\delta t)$ as and the probability of passing to the jth stage is $a_{ij} \lambda_i \delta t + o(\delta t)$, where
$$d_i + \Sigma_j a_{ij} = 1 \quad (i = 1, \ldots, k).$$
Let $p^n(r_1, \ldots, r_k)$ be the probability that death occurs after $n = (r_1 + \ldots + r_k)$ steps, r_i of which have been spent in the ith stage. Then the Laplace transform of the distribution of total life-times is clearly

$$f^*(s) = d_0 + \sum_{r_1, \ldots, r_k} p^n(r_1, \ldots, r_k) \zeta_1^{r_1} \ldots \zeta_k^{r_k}, \tag{8}$$

where $\zeta_i = \lambda_i/(s+\lambda_i)$. Let $p^n(j; r_1, \ldots, r_k)$ be the probability that after $n = (r_1 + \ldots + r_k)$ steps the jth stage is occupied, the ith stage having been occupied r_i times in all. Then

$$p^n(r_1, \ldots, r_k) = \sum_j d_j p^n(j; r_1, \ldots, r_k) \tag{9}$$

and
$$p^n(j; r_1, \ldots, r_k) = \sum_i a_{ij} p^{n-1}(i; r_1, \ldots, r_{j-1}, r_j - 1, r_{j+1}, \ldots, r_k). \tag{10}$$

Write
$$f_j^*(s) = \sum_{r_1, \ldots, r_k} p^n(j; r_1, \ldots, r_k) \zeta_1^{r_1} \ldots \zeta_k^{r_k}, \tag{11}$$

and let m_i, d_0 be the probability of entering the ith stage or of dying at $t = 0$. Then from (10)
$$f_j^*(s) = \sum_i a_{ij} \zeta_j f_i^*(s) + \zeta_j m_j, \tag{12}$$

i.e.
$$(1 + s/\lambda_j) f_j^*(s) - \sum_i a_{ij} f_i^*(s) = m_j. \tag{13}$$

It follows from (13) that $f_j^*(s)$ is the ratio of polynomials in s with the denominator independent of j and of degree k. Finally, since

$$f^*(s) = d_0 + \sum_j d_j f_j^*(s), \tag{14}$$

it follows that $f^*(s)$, too, is of this form. Hence the distribution corresponding to general $a_{ij}, \lambda_i, d_i, m_i$ can be represented in the form (7), though with changed λ_i, and so nothing more general than (7) need be considered.

4. Two further generalizations are possible. We can consider processes in discrete time (5), when the basic distribution is the geometric instead of the exponential. We can also deal with correlated random variables. Suppose that it is required to represent the simplest case of two particles born together at $t = 0$ and having life-times identically, but not independently, distributed. Go back to Erlang's original model with k stages, each with the same λ, and suppose that the particles stay together for the first r stages, and then pass independently through the remaining $(k-r)$ stages. Then the joint

distribution of life-times is (apart from the factor λ) a bivariate χ^2 distribution with $(2k, 2k)$ degrees of freedom and correlation coefficient r/k. The two-dimensional Laplace transform of the distribution is $\lambda^k(\lambda+s_1+s_2)^{-r}(\lambda+s_1)^{r-k}(\lambda+s_2)^{r-k}$. These generalizations will not be considered further in this paper.

5. It was shown in §§ 2 and 3 that by allowing life to have k stages we can deal with distributions of life-time for which $f^*(s)$ is rational with k poles. Call this class of distributions $R_+(k)$; it, or a subclass of it, has been used in a different way in previous analyses of queueing problems (11), (12).

If we allow k to be enumerably infinite we can include those meromorphic $f^*(s)$ for which the appropriate expansion in partial fractions can be made (13). If we allow formal passages to the limit we can include

$$\lim_{k\to\infty}\left(1+\frac{s}{k\lambda}\right)^{-k} = e^{-s/\lambda},$$

representing a constant life-time of λ^{-1}. Erlang did this. Call the whole class of distributions included in this way R_+.

All this raises at least six problems:

(i) What can we say about the whole class R_+ and the subclass $R_+(k)$?

(ii) When is a given rational (or meromorphic) function in $\mathscr{T}_+$?

(iii) Given a distribution of life-time, how can we approximate to it in R_+?

(iv) In applying the device to complicated stochastic systems it will often happen that a solution is only practicable in simple cases. Hence it is worth examining in detail the distributions with small k, with nearly all the λ's equal, and so on.

(v) What is the justification for the formal use of complex probabilities?

(vi) Mr D. G. Kendall has pointed out the importance of defining the class R_+ precisely and of examining its closure properties.

Some of these questions are probably difficult; only a brief discussion will be attempted here.

6. The class $R_+(k)$ for all finite k contains, with any two distributions in it, their convolution and their weighted averages. All distributions in the class are continuous, except possibly for point frequencies at the origin, and the p.d.f. is positive everywhere in $(0, \infty)$ save at isolated points. In the class R_+ there are the constant distribution and hence all discrete distributions,† all distributions in R_+ shifted an arbitrary amount to the right, and all mixtures of a continuous distribution in R_+ with an arbitrary discrete distribution.

Kendall (8) used Schwarz's inequality to prove that for the distributions (3) with real λ_i the coefficient of variation is between 1 and $1/\sqrt{k}$; he was concerned with an application in which the stages have biological significance and in which only $f^*(s)$ of the form (3) need be considered. When we pass to more general $f^*(s)$ the inequalities cease to apply. Thus for $k = 2$ the distribution with Laplace transform

$$(1+s/\mu)(1+s/\lambda_1)^{-1}(1+s/\lambda_2)^{-1}$$

has as $\lambda_2 \to \infty$ and $\mu/\lambda_1 \to 1-$ an arbitrarily large coefficient of variation. Again with $k = 3$ the distribution (5) with $a = b$ has coefficient of variation $\frac{1}{2}$, less than $1/\sqrt{k}$;

† Of course this generality is quite valueless in practice.

$\frac{1}{2}$ is probably the minimum coefficient of variation for $k = 3$. It is very likely that for each k there is a lower bound to the coefficient of variation less than $1/\sqrt{k}$, but I have been unable to estimate how this lower bound changes with k.

7. In a series of papers (see, for example, (9), (10)) Lukacs and Szász† have investigated necessary conditions for a rational function to be in $\mathscr{T}_+$ and have produced a number of special sets of rational functions which are in $\mathscr{T}_+$; (4) is a special case of one of their forms. One simple necessary condition they give can be derived from a general theorem on Laplace transforms which implies that one of the poles with greatest real part must be real. No manageable sufficient condition is known.

8. Consider now the approximation to a given distribution by a distribution in R_+. Since R_+ contains all discrete distributions it is possible in principle to approximate very closely to an arbitrary distribution, but in practice we are usually concerned with small k or with approximations of some simple form. There is an extensive literature (14) on the approximation to functions by rational functions, but it mainly supposes the poles of the rational function to be preassigned, and also is not concerned with rational functions that are in $\mathscr{T}_+$.

If the frequency curve of life-time is given numerically the best procedure is probably to fit linear combinations of exponentials $\Sigma u_i \lambda_i e^{-\lambda_i t}$, with $\Sigma u_i = 1$, by Prony's method (15). Some guidance on the number of terms to include could be obtained from the coefficient of variation of the distribution; sufficient terms would have to be included to make the fitted curve positive in $(0, \infty)$.

If the distribution, $f(t)$ is given analytically it is difficult to give a general procedure. For example, suppose that $f^*(s)$ has a branch point, e.g. $f^*(s) = (1+s/\alpha)^{-\frac{1}{2}}$, corresponding to the χ^2 distribution with one degree of freedom. Since the coefficient of variation is in this case greater than $1/\sqrt{2}$, we can choose a two-stage distribution with the same mean and coefficient of variation as $f(t)$ and satisfying some other requirement such as making the initial ordinate as large as possible; with $k = 3$ a further condition can be satisfied, and so on. The further conditions cannot, however, be taken arbitrarily. This problem will not be investigated further here.

9. Examine now some special cases. If $k = 2$ there can, by the theorem mentioned in §7, be no complex poles. Hence the most general Laplace transform corresponding to a continuous distribution is $f_2^*(s)$, where

$$f_2^*(s) = \frac{(1+s/\mu)}{(1+s/\lambda_1)(1+s/\lambda_2)}, \tag{15}$$

or in the degenerate case

$$\frac{(1+s/\mu)}{(1+s/\lambda)^2}. \tag{16}$$

It is easy to show that (15) is in $\mathscr{T}_+$ if and only if $\mu \geqslant \min(\lambda_1, \lambda_2)$, with a corresponding result for (16). The distribution is unimodal if $\mu > \lambda_1 + \lambda_2$ and J-shaped otherwise; there is a non-zero ordinate at $t = 0$, except when $\mu = \infty$ in (16). An arbitrary discrete probability at $t = 0$ can be included. The coefficient of variation is not less than $1/\sqrt{2}$.

† They did not confine themselves to non-negative variables.

With $k = 3$ a complete enumeration would be very tedious. If the poles are real a series of distributions is obtained with coefficient of variation not less that $1/\sqrt{3}$. With complex poles we can represent any p.d.f. of the form

$$A_1 e^{-\mu_1 t} + A_2 e^{-\mu_2 t} \cos(\alpha t + \beta); \tag{17}$$

one application might be to the representation of a distribution of service-times in a queueing problem, in which there are several different types of servicing operation, a frequently occurring one with a low mean, a less frequent one with a higher mean, and so on. A simple rough approximation to the resulting distribution can be obtained from (17).

For larger k the general nature of the distributions that can be covered is clear, although a precise enumeration of them is difficult. To represent a distribution of coefficient of variation C it is probable that approximately $1/C^2$ stages are required, although it is just possible that a suitable combination of complex poles would allow a representation in appreciably fewer stages. The main gain for small k from the generalization given in this paper is therefore the inclusion of the distributions with trigonometric terms, and an increased flexibility in the representation of the exponential-type distributions; it is disappointing that no great reduction is achieved in the number of stages required to produce a unimodal p.d.f. of small coefficient of variation.

10. Finally, consider the justification of the use of complex probabilities. Bartlett (1) has shown how to construct a formal system including negative probabilities, in which probabilities are manipulated according to the usual rules but in which only positive probabilities have physical significance. Many of his remarks apply here; we are making a subdivision of life into fictitious stages and are not directly concerned with the probabilities corresponding to these stages.

It is, however, worth having a more explicit justification, and one can be obtained by considering a rather general class of processes, with discrete states, which are made into Markov processes where the values of a number of continuous variables, such as expended life-times, are specified. Erlang's method can be shown to be a convenient operational device for solving the integro-differential equation of such processes and to be valid for complex transition probabilities. Since the detailed analysis is of some intrinsic interest it is postponed to a separate paper.

REFERENCES

(1) Bartlett, M. S. *Proc. Camb. phil. Soc.* 41 (1945), 71.
(2) Benson, F. and Cox, D. R. *J.R. statist. Soc.* B, 13 (1951), 65.
(3) Brockmeyer, E., Halstrøm, H. L. and Jensen, A. *The life and works of A. K. Erlang* (Copenhagen, 1948).
(4) Feller, W. *An introduction to probability theory and its applications* (New York, 1950).
(5) Jensen, A. *A distribution model* (Copenhagen, 1954).
(6) Kendall, D. G. *Biometrika*, 35 (1948), 316.
(7) Kendall, D. G. *Ann. Inst. Poincaré*, 13 (1952), 43.
(8) Kendall, D. G. *J.R. statist. Soc.* B, 14 (1952), 41.
(9) Lukacs, E. and Szász, O. *J. Res. nat. Bur. Stand.* 48 (1952), 139.
(10) Lukacs, E. and Szász, O. *J. Res. nat. Bur. Stand.* 52 (1954), 153.
(11) Pollaczek, F. *Math. Z.* 32 (1930), 64 and 729.
(12) Smith, W. L. *Proc. Camb. phil. Soc.* 49 (1953), 449.
(13) Titchmarsh, E. C. *The theory of functions*, 2nd ed. (Oxford, 1939).
(14) Walsh, J. L. *Interpolation and approximation by rational functions in the complex domain* (Colloq. Publ. Amer. math. Soc., New York, 1935).
(15) Whittaker, E. and Robinson, G. *The calculus of observations*, 4th ed. (London, 1944).

Statistical Laboratory
Cambridge