The Annals of Mathematical Statistics
1970, Vol. 41, No. 2, 480–485

INTEGRAL FUNCTIONALS OF BIRTH AND DEATH PROCESSES AND RELATED LIMITING DISTRIBUTIONS[1]

BY DONALD R. MCNEIL

University of Western Australia

1. Introduction. While integrals of nonnegative stochastic processes arise naturally in the theory of inventories and storage (see, for example, Naddor (1966), Moran (1959)), the investigation of moments and distributions of these quantities seems to be a recent development. In an inventory system or storage reservoir the integrated process is of interest because it represents the holding cost associated with the stock in the system over a particular period of time. For queueing processes, particularly those involving automobile traffic such as traffic jams and intersection bottlenecks, the integral of the process up to dissipation is related to the cost of the flow-stopping incident, and it was in this context that Daley (1969), Gaver (1969) and Daley and Jacobs (1969) investigated its distribution. The first of these papers relates to the queue $GI/M/1$, while $M/G/1$ is considered in the other two. In particular, limit theorems as the initial number in the system tends to infinity are established.

Another model in which the integrated process may be important is the classical birth and death process. The growth of insect and other biological populations is often well-described by birth and death processes. Here the integrated process up to extinction has a physical meaning too—it is simply related to the quantity of food consumed, or, in the case of an epidemic, to the number of man-hours lost. In this note we observe and discuss the consequences of the result that the distribution of an integrated birth and death process (or, for that matter, any integral functional of the process) is the same as that of the first passage time for another process with rescaled parameters. We conclude with a discussion of limiting distributions for some special cases.

2. The first passage time. Consider a birth and death process, $X(t), 0 \leqq t < \infty$, defined on the nonnegative integers with initial state $X(0) = i$, and transition probabilities

$$
\begin{aligned}
\Pr[X(t+\delta t) = n+j \mid X(t) = n] &= \mu_n \delta t + o(\delta t), && j = -1 \\
&= 1-(\lambda_n+\mu_n)\delta t + o(\delta t), && j = 0 \\
&= \lambda_n \delta t + o(\delta t), && j = 1 \\
&= o(\delta t), \quad \text{otherwise},
\end{aligned} \tag{1}
$$

where $\lambda_0 \geqq 0$, $\lambda_n > 0$ for $n > 0$, and $\mu_0 = 0$, $\mu_n > 0$ for $n > 0$. If $\lambda_0 = 0$ the process,

Received April 16, 1969.

[1] Research supported by the United States Department of Transportation, Federal Highway Administration, Bureau of Public Roads under Contract FH 11-6884 awarded to the Department of Statistics, The Johns Hopkins University.

having reached the zero state, remains there, but if $\lambda_0 > 0$ transitions out of the zero state are possible. A random variable of interest for such a process is the *first passage time*, Z_i, defined by

$$Z_i = \inf\{t;\ X(t) = 0\}. \tag{2}$$

Of course Z_i may be infinite with non-zero probability. An integral representation for the distribution of Z_i was given by Karlin and McGregor (1957a), and this may be described as follows. Define

$$P_i(t) = \Pr[X(t) = 0 \mid X(0) = i]. \tag{3}$$

Note that if $\lambda_0 = 0$ in (1), we have

$$\Pr[Z_i \leqq x] = P_i(x). \tag{4}$$

In any case it is easily shown that the $P_i(t)$ satisfy

$$P_i'(t) = \mu_i P_{i-1}(t) - (\lambda_i + \mu_i)P_i(t) + \lambda_i P_{i+1}(t). \tag{5}$$

$i = 0, 1, 2, \cdots$, where $P_{-1}(t) = 0$. Equations (5) are the *backward equations* of the birth and death process, and Karlin and McGregor showed that provided

$$\sum_{n=1}^{\infty}\left(\pi_n + \frac{1}{\lambda_n \pi_n}\right) = \infty \qquad \left(\pi_n = \frac{\lambda_1\lambda_2\cdots\lambda_n}{\mu_2\mu_3,\cdots,\mu_{n+1}}\right), \tag{6}$$

equations (5) have a unique solution, and when $\lambda_0 = 0$ this solution is

$$P_i'(t) = \mu_1 \int_0^\infty e^{-zt} Q_i(z)\varphi(dz), \tag{7}$$

where $Q_i(z)$ are polynomials satisfying

$$-xQ_n(x) = \mu_{n+1}Q_{n-1}(x) - (\lambda_{n+1} + \mu_{n+1})Q_n(x) + \lambda_{n+1}Q_{n+1}(x), \tag{8}$$

$(Q_0(x) \equiv 0,\ Q_1(x) \equiv 1)$ and φ is a (unique) measure such that

$$\pi_n \int_0^\infty Q_n(x)Q_m(x)\varphi(dx) = \delta_n^{\ m}. \tag{9}$$

It follows from (4) that Z_i has a probability density function which is given by the right-hand side of equation (7).

While the representation (7) gives immediate insight into the nature of the process, the measure φ is difficult to obtain, in general. Karlin and McGregor (1958a, 1958b, 1959, 1962) evaluated φ for various special cases, in particular when λ_n and μ_n are linear functions of n. In some cases the right-hand side of equation (7) takes the same form as the result obtained using simple generating function techniques (see, for example, Kendall (1948), Bailey (1954)).

3. An integral functional of the process. For a birth and death process $X(t)$ with first passage time Z_i define

$$W_i = \int_0^{Z_i} g\{X(t)\}\,dt, \tag{10}$$

where g is an arbitrary function defined on the positive integers. Put

$$W_i(\theta) = E[e^{-\theta W_i}]. \tag{11}$$

Let S denote the time which elapses before the first transition. Then if A denotes the event that the first transition is a birth and B the complementary event, we have

$$W_i(\theta) = E[e^{-\theta W_i} \mid A] \Pr[A] + E[e^{-\theta W_i} \mid B] \Pr[B],$$
$$= E[e^{-\theta\{W_{i+1}+Sg(i)\}}] \Pr[A] + E[e^{-\theta\{W_{i-1}+Sg(i)\}}] \Pr[B].$$

Using equation (1), this becomes

$$W_i(\theta) = W_{i+1}(\theta) \int_0^\infty e^{-\theta s g(i)} e^{-(\lambda_i+\mu_i)s} \lambda_i \, ds + W_{i-1}(\theta) \int_0^\infty e^{-\theta s g(i)} e^{-(\lambda_i+\mu_i)s} \mu_i \, ds,$$
$$= \lambda_i \{\lambda_i + \mu_i + \theta g(i)\}^{-1} W_{i+1}(\theta) + \mu_i \{\lambda_i + \mu_i + \theta g(i)\}^{-1} W_{i-1}(\theta).$$

Hence

$$\theta W_i(\theta) = \mu_i^* W_{i-1}(\theta) - (\lambda_i^* + \mu_i^*) W_i(\theta) + \lambda_i^* W_{i+1}(\theta), \tag{12}$$

($W_0(\theta) \equiv 1$) where $\lambda_i^* = \lambda_i/g(i)$, $\mu_i^* = \mu_i/g(i)$. But if we take Laplace transforms in equation (5) we get

$$\theta Z_i(\theta) = \mu_i Z_{i-1}(\theta) - (\lambda_i + \mu_i) Z_i(\theta) + \lambda_i Z_{i+1}(\theta), \tag{13}$$

($Z_0(\theta) \equiv 1$) where

$$Z_i(\theta) = \int_0^\infty e^{-\theta t} P_i'(t) \, dt \, (= E[e^{-\theta Z_i}]). \tag{14}$$

Hence equations (12) are essentially nothing other than the backward equations for the birth and death process with rescaled parameters $\lambda_n^* = \lambda_n/g(n)$ and $\mu_n^* = \mu_n/g(n)$, and consequently the distribution of W_i is known whenever the distribution of Z_i is known for the process with parameters λ_n^* and μ_n^*.

Note that if $g(x) \equiv 1$ in equation (10), $W_i = Z_i$, while if $g(x) = x$, W_i is just the area under $X(t)$ up to the time when the process vanishes for the first time. If the process is a queue then Z_1 is the busy period and W_1 is the total waiting time of all customers served during a busy period. For the queue $M/M/1$ (which corresponds to $\lambda_n = \lambda$, $\mu_n = \mu$, $n > 0$) Daley and Jacobs (1969) found the Laplace–Stieltjes transform of the distribution of W_i as the ratio of two Bessel functions but were unable to invert this transform to obtain the distribution explicitly. The reason for the difficulty is now apparent, because the problem is equivalent to finding the first passage time distribution for a birth and death process with parameters λ/n and μ/n, $n > 0$, and for this model the measure φ appearing in (7) is yet to be evaluated.

4. Applications. If one is interested in making inferences about a stochastic process, it often does not matter what statistic is used, as long as its distribution is known. The above result enables the distribution of an easily computable statistic to be found for any birth and death process for which the *ratio* λ_n/μ_n has a simple form, e.g., when λ_n/μ_n is the ratio of two linear functions of n, whereas tests based on other functionals previously considered (the first passage time, the maximum, the total number of transitions before first emptiness, etc.) require that both λ_n and μ_n be simple functions of n. The following example should illustrate this:

For biological populations confined between two limits, N_1 and N_2, say, it is natural to consider a stochastic logistic process (see Moran (1964)) in which

$$\lambda_n = \lambda n(N_2 - n), \qquad \mu_n = \mu n(n - N_1). \tag{15}$$

It is difficult, however, to find the first passage time (to N_1) distribution for this model, and Prendiville (1949) suggested the modification of (15) to

$$\lambda_n = \lambda(N_2 - n), \qquad \mu_n = \mu(n - N_1), \tag{16}$$

which, with a change of origin, gives rise to the classic Ehrenfest model for which the distribution of Z_i is known. If one does not wish to alter the assumptions (15), one can still proceed if attention is directed to W_i with $g(x) = x$ rather than Z_i. In this case W_i is simply related to the quantity of food consumed by the population, and its distribution is the same as that of Z_i where the parameters are given by (16).

Finally, general expressions for the moments of W_i in terms of the λ_n and μ_n are found by using the results of Karlin and McGregor (1957b) for moments of first passage times. In particular

$$E[W_i] = \frac{1}{\mu_i} \sum_{k=0}^{\infty} g(k+1)\pi_k + \sum_{j=0}^{i-2} \frac{1}{\lambda_{j+1}\pi_j} \sum_{k=j+1}^{\infty} g(k+1)\pi_k, \tag{17}$$

whenever (6) is satisfied.

It is of interest to investigate the limiting distributions of W_i as $i \to \infty$ and some special cases are considered below. Results are obtained for Z_i and W_i with $g(x) = x$. Since most of the results concerning first passage time distributions are essentially known, derivations are not given.

5. Limiting distributions in special cases. (a) $\lambda_n = \lambda$, $\mu_n = \mu$, *for all* $n > 0$. The process with these parameters corresponds to the queue $M/M/1$. The distribution of Z_i has (see Bailey (1954)) Laplace–Stieltjes transform $Z_i(\theta) = (\frac{1}{2}\lambda^{-1}\{\lambda+\mu+\theta-[(\lambda+\mu+\theta)^2-4\lambda\mu]^{\frac{1}{2}}\})^i$. When $g(x) = x$, W_i is the total wait in a busy period starting with i customers, and Daley and Jacobs (1969) showed that

$$W_i(\theta) = (\mu/\lambda)^{\frac{1}{2}i} J_{i+(\lambda+\mu)/\theta}\{2(\lambda\mu)^{\frac{1}{2}}/\theta\}/J_{(\lambda+\mu)/\theta}\{2(\lambda\mu)^{\frac{1}{2}}/\theta\},$$

$J_x(z)$ being the Bessel function of order x and argument z. In fact (see Gaver (1969)) the moments are more easily calculated by taking derivatives of (12) at $\theta = 0$ and using generating functions. It is found that provided $\rho < 1$, where $\rho = \lambda/\mu$,

$$E[Z_i] = \frac{i}{\mu(1-\rho)}, \qquad \operatorname{Var}[Z_i] = \frac{i(1+\rho)}{\mu^2(1-\rho)^3}; \qquad E[W_i] = \frac{i^2(1-\rho)+i(1+\rho)}{2\mu(1-\rho)^2},$$

$$\operatorname{Var}[W_i] = \frac{2i^3(1-\rho)^2(1+\rho)+3i^2(1-\rho)(1+6\rho+\rho^2)+i(1+\rho)(1+28\rho+\rho^2)}{6\mu^2(1-\rho)^5}.$$

The asymptotic normality of $\{Z_i - i\mu^{-1}/(1-\rho)\}/i^{\frac{1}{2}}$ follows from the fact that Z_i can be written a sum of independent random variables, each distributed as Z_1. Daley and Jacobs proved that (normalized) W_i also has a limiting normal distribution.

(b) $\lambda_n = \lambda n$, $\mu_n = \mu n$. This process is the linear growth model investigated by Kendall (1948), who derived

$$\Pr[Z_i \leqq x] = \left\{\frac{1-\exp[-\mu(1-\rho)x]}{1-\rho\exp[-\mu(1-\rho)x]}\right\}^i,$$

where $\rho = \lambda/\mu$. On the other hand W_i has the same distribution as Z_i has in example (a), as observed by Karlin (1968, page 335), and this is (Bailey (1954))

$$\Pr[W_i \leqq x] = i\rho^{-\frac{1}{2}i}\int_0^x e^{-\mu(1+\rho)t} I_i(2\mu\rho^{\frac{1}{2}}t)\,dt,$$

where $I_n(z)$ is the modified Bessel function of order n and argument z. Note that

$$\Pr\left[Z_i - \frac{\log i}{\mu(1-\rho)} \leqq x\right] = \left\{\frac{1-(1/i)\exp[-\mu(1-\rho)x]}{1-(\rho/i)\exp[-\mu(1-\rho)x]}\right\}^i \to \exp[-(1-\rho)e^{-\mu(1-\rho)x}]$$

as $i \to \infty$. Thus, if $\rho < 1$, $Z_i - (\log i)/\{\mu(1-\rho)\}$ has a limiting extreme value (Gompertz) distribution as $i \to \infty$, with location and scale parameters $\{\log(1-\rho)\}/\{\mu(1-\rho)\}$ and $\mu(1-\rho)$ respectively.

(c) $\lambda_n = \lambda$, $\mu_n = \mu n$. In this case the process corresponds to the queue $M/M/\infty$, or, alternatively, the immigration-death process. The distribution of Z_i was obtained by Karlin and McGregor (1958a), and is

$$\Pr[Z_i \leqq x] = 1 - \mu \sum_{n=0}^{\infty} \frac{\alpha_n}{s_n} c_i^*(s_n/\mu;\, \lambda/\mu) e^{-s_n x}, \tag{18}$$

where $s_0 < s_1 < s_2 < \cdots (\mu n < s_n < \mu n + \mu)$ are the (simple) zeros of

$$B(s) = e^{-\lambda/\mu} \sum_{n=0}^{\infty} \frac{(\lambda/\mu)^n}{n!(n\mu - s)},$$

$\alpha_n = \{\lambda\mu B'(s_n)\}^{-1}$ and $c_i^*(x;a)$ are polynomials satisfying the recurrence relations $ac_i^*(x;a) + (x-i-a)c_{i-1}^*(x;a) + ic_{i-2}^*(x;a) = 0$, $(c_{-1}^*(x;a) \equiv 0,\ c_0^*(x;a) \equiv 1)$.

To discuss the limiting situation as $i \to \infty$, write

$$Z_i = G_i + G_{i-1} + \cdots + G_1, \tag{19}$$

where G_j is the first passage time from state j to state $j-1$. It is clear that the G_j are independent, and for large j,

$$E[G_j] \sim \frac{1}{\mu j}, \qquad \mathrm{Var}[G_j] \sim \left(\frac{1}{\mu j}\right)^2.$$

It follows that random variables $G_j - E[G_j]$ do not satisfy the Lindeberg condition for asymptotic normality (see Feller (1966) page 256), but $Z_i - E[Z_i]$ does have a limiting distribution with zero mean and finite variance as $i \to \infty$ (Feller page, 259). This distribution has the same form as that of $Z_i - \mu^{-1}\log i$, and, using equation (18), this is

$$\Pr[Z_i - \mu^{-1}\log i \leqq x] = 1 - \mu \sum_{n=0}^{\infty} \frac{\alpha_n}{s_n}\left(\frac{1}{i}\right)^{s_n/\mu} c_i^*(s_n/\mu;\, \lambda/\mu) e^{-s_n x}. \tag{20}$$

The limiting distribution is thus obtained by taking the limit as $i \to \infty$ of the right-hand side of equation (20).

As $i \to \infty$ the polynomials $c_i^*(x; a)$ have the same asymptotic form as the classical Charlier polynomials (Erdélyi (1953) page 226) and thus $(i)^{-x}c_i^*(x; a) \to (-a)^{-x}$ as $i \to \infty$.

Consequently equation (20) becomes, in the limit as $i \to \infty$,

$$\lim_{i\to\infty} \Pr[Z_i - \mu^{-1}\log i \leqq x] = 1 - \mu \sum_{n=0}^{\infty} \frac{\alpha_n}{s_n}\left(\frac{\lambda}{\mu}\right)^{s_n/\mu} \cos(\pi s_n/\mu) e^{-s_n x},$$

that is, the limiting distribution of $Z_i - E[Z_i]$ has an exponential tail.

The limiting distribution of W_i is much easier to obtain. Writing (as in equation (19)) $W_i = H_i + H_{i-1} + \cdots + H_1$, it is easily shown that $H_j(\theta) = E[e^{-\theta H_j}]$ satisfies

$$(\lambda/j + \mu + \theta)H_j(\theta) = (\lambda/j)H_{j+1}(\theta)H_j(\theta) + \mu, \tag{21}$$

so $H_j(\theta) \sim \mu/(\mu+\theta)$ as $j \to \infty$. Consequently the distribution of $(\mu/i^{\frac{1}{2}})\{W_i - E[W_i]\}$ is asymptotically standardized normal, using Lyapunov's theorem (Feller, page 278).

REFERENCES

[1] Bailey, N. T. J. (1954). A continuous time treatment of a simple queue using generating functions. *J. Roy. Statist. Soc. Ser. B* **16** 288–291.

[2] Daley, D. J. (1969). The total waiting time in a busy period of a stable single-server queue, I. (To appear in *J. Appl. Probability*)

[3] Daley, D. J. and Jacobs, D. R., Jr. (1969). The total waiting time in a busy period of a stable single-server queue, II. (To appear in *J. Appl. Probability*.)

[4] Erdélyi, A. et al. (1953). *Higher Transcendental Functions*, **2**. McGraw-Hill, New York.

[5] Feller, W. (1966). *An Introduction to Probability Theory and Its Applications*, **2**. Wiley, New York.

[6] Gaver, D. P. (1969). Highway delays resulting from flow-stopping incidents, *J. Appl. Probability*, **6** 137–153.

[7] Karlin, S. (1968). *A First Course in Stochastic Processes*. Academic Press, New York.

[8] Karlin, S. and McGregor, J. (1957a). The differential equations of birth and death processes, and the Stieltjes moment problem. *Trans. Amer. Math. Soc.* **85** 489, 546.

[9] Karlin, S. and McGregor, J. (1957b). The classification of birth and death processes. *Trans. Amer. Math. Soc.* **86** 366–400.

[10] Karlin, S. and McGregor, J. (1958a). Many server queueing processes with Possion input and exponential service times. *Pacific J. Math.* **8** 87–118.

[11] Karlin, S. and McGregor, J. (1958b). Linear growth, birth and death processes. *J. Math. Mech.* **7** 643–662.

[12] Karlin, S. and McGrfgor, J. (1959). Random walks. *Illinois J. Math.* **3** 66–81.

[13] Karlin, S. and McGregor, J. (1962). On a genetics model of Moran. *Proc. Cambridge Philos. Soc.* **58** 299–311.

[14] Kendall, D. G. (1948). On the generalized "Birth and Death" process. *Ann. Math. Statist.* **19** 1–15.

[15] Moran, P. A. P. (1959). *The Theory of Storage*. Methuen, London.

[16] Moran, P. A. P. (1964). Some general results on random walks, with genetic application. *J. Australian Math. Soc.* **3** 469–479.

[17] Naddor, E. (1966). *Inventory Systems*. Wiley, New York.

[18] Prendiville, B. J. (1949). Discussion of paper by Kendall. *J. Roy. Statist. Soc. Ser. B* **11** 273.