

ON A FORMULA IN DIFFUSION PROCESSES IN POPULATION GENETICS

CHARLES J. HOLLAND

ABSTRACT. A problem of recent interest has been the determination of the average values of certain quantities conditioned upon the fixation (or extinction) of the gene in the population. In this note we give another simple proof of formulas for the determination of these quantities.

In this note we give a simple derivation of a formula which has recently been of importance in the study of population genetics using diffusion methods. Nagylaki [5] has recently given another derivation using different methods. Suppose that the gene frequency $\xi_x(t, \omega)$ evolves over the interval $[0, 1]$ according to a diffusion process with initial constant frequency $\xi_x(0, \omega) = x$, mean f , and covariance σ^2 where f, σ are uniformly Lipschitz on every compact subinterval of $[0, 1]$. Henceforth we drop notational dependence upon the path ω . Let τ_x denote the first time $t \geq 0$ such that $\xi_x(t) = 0$ or 1, then the random variable τ_x represents the extinction or fixation time depending upon whether $\xi_x(\tau_x) = 0$ or 1 respectively. For $t \leq \tau_x$, $\xi_x(t)$ satisfies the scalar Ito stochastic differential equation

$$(1) \quad d\xi_x(t) = f(\xi_x(t)) dt + \sigma(\xi_x(t)) dW(t)$$

where W is one dimensional brownian motion. For L a continuous function on $[0, 1]$ and j a positive integer the quantities of interest are

$$(2) \quad Q_j(x) = E \left\{ \left[\int_0^{\tau_x} L(\xi_x(t)) dt \right]^j \mid \xi_x(\tau_x) = 1 \right\},$$

$$(3) \quad R_j(x) = E \left\{ \left[\int_0^{\tau_x} L(\xi_x(t)) dt \right]^j \mid \xi_x(\tau_x) = 0 \right\}.$$

See Ewens [2] and Maruyama and Kimura [4] for their applications in population genetics.

In Theorem 1 we verify the correctness of the equations (4) below given in [4] for the determination of $Q_j(x)$. The quantities $R_j(x)$ are determined in a similar manner. Define the differential operator

$$A = \frac{1}{2} \sigma^2(x) \frac{d^2}{dx^2} + f(x) \frac{d}{dx}.$$

THEOREM 1. *Suppose for some positive integer k that $E[\tau_x^k] < \infty$ and that T_j , $j = 0, 1, \dots, k$, are solutions with continuous second derivatives on $[0, 1]$ to the equations*

Received by the editors March 24, 1975.

AMS (MOS) subject classifications (1970). Primary 60J70; Secondary 92A10.

© American Mathematical Society 1976

$$(4) \quad AT_j(x) + jL(x)T_{j-1}(x) = 0$$

with $T_{-1} \equiv 0$ and boundary conditions $T_0(0) = 0, T_0(1) = 1$, and $T_j(0) = T_j(1) = 0, j = 1, \dots, k$. Then

$$(5) \quad T_0(x) = \text{Prob}\{\xi_x(\tau_x) = 1\},$$

$$(6) \quad Q_j(x) = T_j(x)/T_0(x) \text{ for } 0 < x \leq 1, \quad j = 1, 2, \dots, k.$$

PROOF. Since $E[\tau_x] < \infty$, then $\tau_x < \infty$ with probability one and (5) follows from the Ito stochastic differential rule; see [3, p. 108]. The assumption $\sigma^2 > 0$ there is used only to guarantee the existence of a solution to (4) with $j = 0$. For any positive integer $r \leq k$, define $y_0 = 1, y = (y_1, \dots, y_r)$ and

$$(7) \quad \phi(x, y) = \sum_{j=0}^r \binom{r}{j} T_{r-j}(x) y_j.$$

Note that ϕ satisfies the equation

$$(8) \quad \frac{1}{2} \sigma^2 \phi_{xx} + f \phi_x + \sum_{j=1}^r jL y_{j-1} \phi_{y_j} = 0.$$

Consider the equations (1) and

$$(9) \quad d\eta_j(t) = jL(\xi_x(t))\eta_{j-1}(t) dt, \quad \eta_j(0) = 0,$$

for $j = 1, \dots, r$ and $\eta_0 \equiv 1$. It is easy to see that for $t \leq \tau_x, \eta_j(t) = [\int_0^t L(\xi_x(s)) ds]^j$.

Let $\bar{\tau}_x = \min(t, \tau_x)$, then the Ito stochastic differential rule yields

$$(10) \quad \phi(x, 0) = E \{ \phi(\xi_x(\bar{\tau}_x), \eta_1(\bar{\tau}_x), \dots, \eta_r(\bar{\tau}_x)) \}.$$

Since $E\{\tau_x^k\} < \infty$ and L is bounded on $[0, 1]$, then for $j = 0, 1, 2, \dots, r - 1$,

$$(11) \quad \lim_{t \rightarrow \infty} E \{ T_{r-j}(\xi_x(\bar{\tau}_x)) \eta_j(\bar{\tau}_x) \} = 0.$$

The last equality follows from the facts that $\xi_x(\tau_x) = 0$ or 1 and the boundary condition $T_j(0) = T_j(1) = 0$. Let $\chi = 1$ if $\xi_x(\tau_x) = 1, \chi = 0$ if $\xi_x(\tau_x) = 0$. Using (7), (10), (11) and the boundary conditions on T_0 , we have

$$T_r(x) = \phi(x, 0) = E \left\{ \left[\int_0^{\tau_x} L(\xi_x(t)) dt \right]^r \cdot [\chi = 1] \right\}.$$

It follows immediately from the Bayes rule definition of conditional expectation that

$$Q_r(x) = T_r(x)/T_0(x).$$

This proves Theorem 1.

Now suppose that $\bar{T}_j(x), j = 1, \dots, k$, are solutions to

$$(12) \quad A\bar{T}_j(x) + jL(x)\bar{T}_{j-1}(x) = 0,$$

$\bar{T}_0(x) \equiv 1$, satisfying the boundary conditions $\bar{T}_j(0) = \bar{T}_j(1) = 0$. If $E\{\tau_x^k\} < \infty$, the same proof as in Theorem 1 shows that

$$\bar{T}_j(x) = E \left\{ \int_0^{\tau_x} L(\xi_x(t)) dt \right\}^j, \quad j \leq k.$$

For $\sigma^2 > 0$ on $[0, 1]$, this result is contained in Theorem 13.17 in [1].

Let us now discuss the assumption $E\{\tau_x^k\} < \infty$. If $\sigma^2 > 0$ on $[0, 1]$, then $E\{\tau_x^k\} < \infty$ for any positive integer k . Unfortunately $\sigma^2(0) = \sigma^2(1) = 0$ for most models in population genetics. However, we have the following result which replaces the assumption $E\{\tau_x^k\} < \infty$.

THEOREM 2. *Suppose $S_j, j = 1, 2, \dots, k$, are solutions with continuous second derivatives to $AS_j(x) + jS_{j-1}(x) = 0, S_j(0) = S_j(1) = 0$, with $S_0(x) \equiv 1$. Then $S_j(x) = E[\tau_x^j] < \infty$.*

PROOF. We establish the result by induction. For S_1 see [3, p. 108]. The assumption $\sigma^2 > 0$ on $[0, 1]$ is used there only to guarantee the existence of the solution for S_1 . Suppose the result is true for $j = 1, 2, \dots, r - 1$ with $r \leq k$. Then we show the result is true for $j = r$. Define $y_0 = 1, y = (y_1, \dots, y_r)$ and

$$\phi(x, y) = \sum_{j=0}^r \binom{r}{j} S_{r-j}(x) y_j.$$

Note that ϕ satisfies (8) with $L = 1$. Consider now the equations (9) with $L = 1$. Then

$$S_r(x) = E \sum_{j=0}^r \binom{r}{j} S_{r-j}(\xi_x(\bar{\tau}_x)) (\bar{\tau}_x)^j$$

where $\bar{\tau}_x$ is as defined in the proof of Theorem 1. Using the induction hypothesis and the Lebesgue dominated convergence theorem one obtains

$$\lim_{t \rightarrow \infty} E \binom{r}{j} S_{r-j}(\xi_x(\bar{\tau}_x)) (\bar{\tau}_x)^j = 0$$

for $j = 0, 1, \dots, r - 1$. Hence $S_r(x) = \lim_{t \rightarrow \infty} E(\bar{\tau}_x^r)$ and since $\bar{\tau}_x$ increases to τ_x , then $S_r(x) = E\{\tau_x^r\}$. This proves Theorem 2.

REFERENCES

1. E. B. Dynkin, *Markov processes*, Fizmatgiz, Moscow, 1963; English transl., Vol. II, Die Grundlehren der math. Wissenschaften, Band 122, Academic Press, New York; Springer-Verlag, Berlin, 1965. MR 33 #1887.
2. W. J. Ewens, *Conditional diffusion processes in population genetics*, Theoret. Population Biology 4(1973), 21-30.
3. I. I. Gihman and A. V. Skorohod, *Stochastic differential equations*, "Naukova Dumka", Kiev, 1968; English transl., Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 72, Springer-Verlag, Berlin and New York, 1972. MR 41 #7777.
4. T. Maruyama and M. Kimura, *Some methods for treating continuous stochastic processes in population genetics*, Japan. J. Genetics 46(1971), 407-410.
5. T. Nagylaki, *The moments of stochastic integrals and the distribution of sojourn times*, Proc. Nat. Acad. Sci. U.S.A. 71(1974), 746-749.