

**STOCHASTIC INTEGRALS AND THEIR RELATION TO SOME DIFFUSION MODELS  
OF POPULATION GROWTH AND GENE FREQUENCY**

**Henry C. Tuckwell and John B. Walsh**

**Department of Mathematics University of British Columbia**

**Vancouver, B.C., Canada**

## Abstract

The Ito and Stratonovich stochastic integrals and their relevance to modelling the random growth of populations are discussed. It is seen that both types of integrals may arise naturally in certain circumstances. Three related diffusion models for population growth are derived from discrete models, and are applied to model the behaviour of gene frequency in a haploid population.

### 1. Introduction

A rule of thumb of ordinary calculus is that one can ignore second order terms. That is,  $(\Delta X)^2$  is insignificant compared to  $\Delta X$ . In contrast, the corresponding rule of stochastic calculus is that one can ignore third order terms but that second order terms are significant. In fact, if  $\{W_t, t \geq 0\}$  is a standard one-dimensional Brownian motion (Wiener process), then the quantity

$$Q_n(t) = \sum_{k=0}^{n-1} \left( W_{\frac{(k+1)t}{n}} - W_{\frac{kt}{n}} \right)^2$$

converges almost surely to  $Q(t) = t$  as  $n \rightarrow \infty$ . This property is at the heart of much of the subtlety of the subject of stochastic integration, and equally at the base of much of the confusion.

In particular, it is the reason for the existence of several non-equivalent definitions of stochastic integrals. Two

of these will concern us here: the Ito integral (Ito, 1951 a, b; Doob, 1953) and the Stratonovich integral (Gray and Caughey 1965; Stratonovich, 1966). These two integrals differ even in the simplest cases. For instance,  $\int_0^t W_s dW_s$  equals  $\frac{1}{2} W_t^2$  as a Stratonovich integral, whereas it equals  $\frac{1}{2} W_t^2 - \frac{1}{2} t$  as an Ito integral\*.

Each integral has its advantages: the Ito integral is defined for a much larger class of integrands and has better convergence properties, while the Stratonovich integral has a simpler calculus and is easier to use in complex calculations. However, both integrals are adequate to handle the stochastic differential equations encountered in biology and there seems to be no overriding reason for preferring one to the other.

The existence of two stochastic calculi has led to confusion, and certain stochastic differential equations proposed to model biological phenomena have become centers of controversy: should they be interpreted as equations in Ito's calculus or Stratonovich's?

---

\* Both these integrals can be calculated directly. It is instructive to compare the two calculations, for they show the role of the quadratic variation  $Q(t)$  quite clearly (McKean 1969).

Although the first reference in the biological literature to the potential difficulties raised by the non-uniqueness of the stochastic integral was May (1971), the problem had received considerable attention in the context of non-linear random systems in engineering (Gray & Caughey, 1965; Mortensen, 1969).

Such discussions have not resolved the problem faced by the theoretical ecologist, as is evidenced by the controversy over the interpretation of various equations. For example, May (1971) put forward a version of the random logistic growth process

$$(1.1) \quad dN = N(k_0 - N) dt + \sigma N dW$$

where  $\sigma$  is a variance parameter and  $k_0$  is the "mean carrying capacity". May proposed this as an Ito equation, with a view to finding the stationary distribution of  $N_t$ . Noting that May had given no reason for preferring the Ito version to the Stratonovich version of (1.1), Feldman and Roughgarden (1975) pointed out that if one analyzes it as a Stratonovich equation one finds that a non-degenerate stationary distribution always exists, whereas May had shown that the stationary distribution for the Ito version is degenerate unless  $k_0 > \frac{1}{2} \sigma^2$ .

Similar problems have arisen in population genetics. If  $X$  is the frequency of an allele in a haploid population with two

alleles, then, following Crow and Kimura (1970), a deterministic model for the behavior of  $X$  is

$$\frac{dX}{dt} = sX(1-X)$$

where  $s$  is the selection coefficient. Kimura (1954, 1964) then suggested a diffusion model of the same phenomenon. He studied the diffusion directly, using Kolmogorov's forward differential equation rather than stochastic integrals, but one can still describe his diffusion in terms of stochastic differential equations. It satisfies

$$(1.2) \quad dX = sX(1-X) dt + \sigma X(1-X) dW$$

where the differential is taken in the Ito sense. Tuckwell (1976) found conditions for quasifixation when (1.2) is taken in the sense of a Stratonovich differential. Coincidentally, Gillespie (1972) derived a diffusion equation for  $X$ , starting with a discrete model, and found that it satisfies equation (1.2) with the differentials again taken in the Stratonovich sense.

It might seem curious that the question of Ito versus Stratonovich integrals should come up in connection with a model which had been derived completely independently of stochastic

integration, but in fact the non-uniqueness of stochastic integrals is more a symptom than the cause of the difficulties. Diffusion processes, like Brownian motion, have non-zero quadratic variation, and, as with stochastic integrals, one cannot ignore second order terms.

It is tempting to construct models by simply putting some noise into an existing ordinary differential equation or by writing down a reasonable-looking stochastic differential equation, but the fact that there can be disagreement over the simple examples above shows that diffusion models are subtler than they appear. The problem is not the choice of stochastic calculus; it is that there are pitfalls in translating physical assumptions directly into a continuous-parameter model. The intuition developed from working with ordinary calculus is not always a reliable guide.

As a practical matter, we agree with Ludwig (1976) that one should start with a discrete model if one is available, for it is much easier to specify the hypotheses clearly - and in particular to see what independence relations there should be between the different

random variables - in a discrete model than in a continuous model. Once this is done, one can find the corresponding diffusion approximation if it exists.

In sections 3 and 4 we will consider several simple diffusion models of population growth and gene frequency which can be derived from underlying discrete models. We will try to show which type of stochastic integral enters and why. In order to illustrate some of the available methods we will take different approaches to each of the models, but we will see later that all the models we consider can be unified by the idea of a branching process in a random environment.

The final section is devoted to a general but brief discussion of the problem of deriving stochastic differential equations.

## 2. Stochastic Integrals

Before passing on to specific equations we will introduce some notation and mention some well-known facts concerning stochastic integrals and differential equations, and the translation from one calculus to another. A more detailed account may be found in Jaswinski (1970). We will restrict our attention to processes which are temporally homogeneous.

One commonly writes stochastic differential equations either in the differential form

$$(2.1) \quad dN = f(N) dt + g(N) dW$$

or in the form

$$(2.2) \quad \frac{dN}{dt} = f(N) + g(N) \dot{W}$$

where  $\dot{W}$  is "white noise". Neither (2.1) nor (2.2) should be regarded as having meaning in itself. Both are merely shorthand for the integral equation

$$(2.3) \quad N_t = N_0 + \int_0^t f(N_s) ds + \int_0^t g(N_s) dW_s.$$

We must of course specify which integral, Ito or Stratonovich, we are using in (2.3), and it follows that neither (2.1) nor (2.2) is well-defined until this is done. Though this is an obvious point, we mention it to counteract the natural impression that the "increments"  $dN$  and  $dW$  are well-defined. They are not: in fact the equation

$$(2.4) \quad dN = NdW$$

has the solution  $N_t = N_0 e^{W_t}$  if the Stratonovich integral is employed,

whereas  $N_t = N_0 e^{W_t - \frac{1}{2}t}$  if one uses the Ito integral.

It is easy to translate integrals and equations from one calculus to another. If  $f(t, x)$  is continuously differentiable in  $x$  and piecewise continuous in  $t$ ,

$$(2.5) \int_0^t f(s, W_s) dW_s \text{ (Stratonovich)} = \frac{1}{2} \int_0^t \frac{\partial f}{\partial x}(s, W_s) ds + \int_0^t f(s, W_s) dW_s \text{ (Ito)}.$$

The basic difference between the two calculi is in the chain rule. If  $F(t, x)$  is continuously differentiable in  $t$  and twice continuously differentiable in  $x$ , then

$$(2.6) dF(t, W_t) = \frac{\partial F}{\partial t}(t, W_t) dt + \frac{\partial F}{\partial x}(t, W_t) dW_t \text{ (Stratonovich)}$$

in the Stratonovich calculus, while in Ito's calculus:

$$(2.7) dF(t, W_t) = \frac{\partial F}{\partial t}(t, W_t) dt + \frac{1}{2} \frac{\partial^2 F}{\partial x^2}(t, W_t) dt + \frac{\partial F}{\partial x}(t, W_t) dW_t \text{ (Ito)}.$$

Furthermore, if  $g$  and  $h$  are twice continuously differentiable, then

$$(2.8) \quad dN = g(N)dt + h(N)dW \quad \text{(Ito)}$$

and

$$(2.9) \quad dN = (g(N) - \frac{1}{2} h'(N) h(N)) dt + h(N) dW \quad (\text{Stratonovich})$$

are the same equation, in the sense that they both have the same solutions.

The solution of (2.8) is a diffusion process and can be studied by its Kolmogorov differential equations, as well as by the theory of stochastic differential equations. Its transition probability density  $p(t, N|N_0)$  will satisfy the (forward) Kolmogorov equation

$$(2.10) \quad \frac{\partial p}{\partial t} = - \frac{\partial}{\partial N} (g(N)p) + 1/2 \frac{\partial^2}{\partial N^2} (h(N)^2 p) .$$

However, the diffusion processes we derive in this paper will be defined by the stochastic differential equations they satisfy rather than by the Kolmogorov equations. This is partly to provide a unified treatment and partly because we will be led in a natural way to processes which are not diffusions, but which do satisfy stochastic differential equations.

### 3. Diffusion Approximations for Malthusian Growth

We now consider three discrete-time models of Malthusian growth. We will derive the stochastic differential equations whose solutions approximate the discrete-time processes. These are not competing models; rather they place emphasis on different aspects of population growth.

Our derivations are heuristic, but they can be made rigorous and it will be clear which type of stochastic integral enters and why. The first model can be tackled using the central limit theorem, which has been employed in this context by various authors (Lewontin & Cohen, 1969; Gillespie, 1972; Cook and Hartl, 1974).

The discrete-time processes we consider converge in distribution to diffusions. Since only the distributions are important, we can replace a given process by another having the same or nearly the same distribution. Hence if  $X_1, X_2, \dots$  is a random walk with  $E\{X_{n+1} - X_n\} = \mu$  and  $\text{Var}\{X_{n+1} - X_n\} = \sigma^2$ , then by the central limit theorem  $X_n$  is nearly normal for large  $n$ , and, moreover, by standard weak-convergence results, the whole process  $X_n, X_{n+1}, X_{n+2}, \dots$  has approximately the same distribution as

$\{(n+1)\mu + \sigma W_{n+1}, i = 0, 1, 2, \dots\}$  for large  $n$ . That is, we can replace  $X_n$  by  $n\mu + \sigma W_n$ .

### A. THE RANDOM ENVIRONMENT MODEL

Let  $Z_0, Z_1, \dots$  be i.i.d random variables and suppose that  $P\{Z_0 \geq \delta\} = 1$  for some  $\delta > 0$ , and that  $E\{Z_0\} < \infty$ . Let  $N_n$  be the population size at time  $n$  and suppose that

$$(3.1) \quad N_{n+1} = Z_n N_n \quad n = 0, 1, 2, \dots$$

The  $Z_i$  represent an environmental effect which affects all members of the population alike, hence it acts multiplicatively. If we set  $\mu = E\{\log Z_0\}$  and  $\sigma^2 = \text{Var}\{\log Z_0\}$ , we will see that the diffusion approximation to (3.1) satisfies the Ito equation

$$(3.2) \quad dN = (\mu + \frac{1}{2}\sigma^2)Ndt + \sigma N dW \quad (\text{Ito}).$$

To see this, note that from (3.1),  $N_n = N_0 \prod_{j=0}^{n-1} Z_j$ .

Let  $S_n = \log N_n$  and  $D_n = \log Z_n$ , so that

$$S_n = S_0 + \sum_{j=0}^{n-1} D_j.$$

The  $D_j$  are i.i.d. so  $\{S_n\}$  is a random walk. Under our hypotheses the  $D_j$  have moments of all orders, hence the central limit theorem applies, and, as we remarked above, if  $Y_n = S_0 + n\mu + \sigma W_n$ , the whole process  $S_n, S_{n+1}, \dots$  has nearly the same distribution as  $Y_n, Y_{n+1}, \dots$ . This means that  $e^{S_n}, e^{S_{n+1}}, \dots$  has nearly the same distribution as  $N_n, N_{n+1}, \dots$ , and  $N_n$  can be approximated by the process

$$(3.3) \quad N_t = N_0 e^{\mu t + \sigma W_t}, \quad t \geq 0.$$

To see what stochastic differential equation  $N_t$  satisfies, we apply Ito's formula (2.7) to obtain

$$(3.4) \quad dN = N_0 e^{\mu t + \sigma W_t} ((\mu + \frac{1}{2} \sigma^2) dt + \sigma dW),$$

which gives (3.2). This is an Ito equation because we used Ito's formula to pass from (3.3) to (3.4). Had we used (2.6), the corresponding formula in the Stratonovich calculus, we would have obtained

$$(3.5) \quad dN = \mu N dt + \sigma N dW \quad (\text{Stratonovich}).$$

This equation is just the translation of (3.2) into the Stratonovich calculus using (2.8) and (2.9).

In the above derivation we were able to first find the solution and then find the equation it satisfies. It was only the extreme simplicity of the model that allowed us to do this. The next model is slightly more complex and we will construct the stochastic integral equation directly.

#### B. THE BRANCHING MODEL

From a slightly different point of view one might suppose that the environment affects each individual independently. If at time  $n$  there are  $N_n$  individuals in the population, then each individual will be replaced by a certain number of offspring in the next generation, independently of the rest of the population.

Thus, let  $V_{nj}$ ,  $n, j = 1, 2, \dots$  be i.i.d. random variables with  $E\{V_{11}\} = Z$  and  $\text{Var}\{V_{11}\} = \sigma^2$ . Then we put

$$(3.6) \quad N_{n+1} = \sum_{j=1}^{N_n} V_{nj},$$

where the sum is the random variable which equals  $\sum_{j=1}^k V_{nj}$  on the set  $\{N_n = k\}$ .  $N_n$  is in fact a branching process (Feller, 1951) and we will show that its diffusion approximation satisfies the Ito equation

$$(3.7) \quad dN = \mu N dt + \hat{\sigma} \sqrt{N} dW \quad (\text{Ito})$$

where  $\mu = \log Z$ .

To see this, let  $Y_{nj} = V_{nj} - Z$  so that

$$N_{n+1} = Z N_n + \sum_{j=1}^{N_n} Y_{nj}.$$

Conditioned on the value of  $N_n$ , the sum above is independent of the past up to time  $n$  and has mean zero and variance  $N_n \sigma^2$ . Furthermore, if  $N_n$  is large, the central limit theorem tells us that the sum is nearly normal. Thus it has nearly the same distribution as  $\hat{\sigma} \sqrt{N_n} (W_{n+1} - W_n)$ , so that, approximately,

$$(3.8) \quad N_{n+1} = Z N_n + \hat{\sigma} \sqrt{N_n} \Delta W,$$

which defines  $\Delta W$ . Now we suppose that successive generations come at shorter and shorter intervals, and look at the process at times  $\epsilon, 2\epsilon, \dots$ . Defining  $\Delta_{n\epsilon} W = W_{(n+1)\epsilon} - W_{n\epsilon}$ , we replace (3.8) by

$$(3.9) \quad N_{(n+1)\epsilon} = Z^\epsilon N_{n\epsilon} + \hat{\sigma} \sqrt{N_{n\epsilon}} \Delta_{n\epsilon} W,$$

the choice of the constant  $Z^\epsilon$  in (3.9) follows from the fact that  $E\{N_{(n+1)\epsilon}\} = Z^\epsilon E\{N_{n\epsilon}\} = Z^{n\epsilon} E\{N_0\}$ . If we take  $n = \frac{1}{\epsilon}$ , this says that  $E\{N_{n\epsilon}\} = Z = E\{N_1\}$ , which agrees with (3.8). Thus

$$(3.10) \quad N_{(n+1)\epsilon} - N_{n\epsilon} = (Z^\epsilon - 1) N_{n\epsilon} + \hat{\sigma} \sqrt{N_{n\epsilon}} \Delta_{n\epsilon} W$$

Summing, we get

$$N_{n\epsilon} = N_0 + \frac{1}{\epsilon} (Z^\epsilon - 1) \sum_{j=1}^{n-1} N_{j\epsilon} \epsilon + \hat{\sigma} \sum_{j=0}^{n-1} \sqrt{N_{j\epsilon}} \Delta_{j\epsilon} W.$$

Now, if we define  $N_t^k$  for non-integer  $t$  by

$N_t^k = N_{j\epsilon}$  if  $j\epsilon \leq t < (j+1)\epsilon$ , we recognize these sums as integrals:

$$(3.11) \quad N_{n\epsilon}^k = N_0 + \frac{1}{\epsilon} (Z^\epsilon - 1) \int_0^{n\epsilon} N_t^k dt + \hat{\sigma} \int_0^{n\epsilon} \sqrt{N_t^k} dW$$

where the last integral is an Ito integral. (This can be seen from

the fact that  $\Delta_{j\epsilon} W = W_{(j+1)\epsilon} - W_{j\epsilon}$ , i.e. the increment sticks out

in the future whereas the integrand is evaluated at the beginning of

the interval.) Now suppose that  $N_t^k$  converges to a limit process  $N_t$ .

Then the integrals on the right-hand side will converge, and in particular the stochastic integral will converge to an Ito integral to give:

$$N_t = N_0 + (\log Z) \int_0^t N_s ds + \hat{\sigma} \int_0^t \sqrt{N_s} dW_s, \quad (\text{Ito})$$

which is just (3.7). Note that the Stratonovich version of (3.7) is

$$(3.12) \quad dN = \left( \mu N + \frac{\hat{\sigma}^2}{4} \right) dt + \hat{\sigma} \sqrt{N} dW.$$

One can complicate the above model slightly by allowing certain age-dependent or multi-type branching processes. One, which leads to a model put forward by Levins (1969), is worth mentioning here. It was analyzed by Tuckwell (1974) and Kiester and Barakat (1974), but its origins were questioned by Keiding (1975). This comes from a two-type branching process in which the two types are offspring and adults. At each generation, the  $i^{\text{th}}$  adult has  $X_i$  offspring, where the  $X_i$  are i.i.d. with mean  $\mu$  and variance  $\sigma^2$ . If an adult has probability  $v$  of surviving to the next breeding season and each offspring has probability  $u$  of surviving to adulthood and the next breeding season, then the total population  $N$  has a diffusion approximation satisfying the Ito equation

$$(3.13) \quad dN = (v+\mu u-1) N dt + \sqrt{v(1-v) + \mu u(1-u) + \sigma^2 u^2} \sqrt{N} dW$$

If we had defined  $X_i$  to be the number of surviving children (or equivalently, if we had set  $u=1$ ) and if the number of offspring is fixed ( $\sigma^2 = 0$ ) we get

$$dN = (v+\mu-1) N dt + \sqrt{v(1-v)} \sqrt{N} dW \quad (\text{Ito}),$$

which was the equation proposed by Levins, with  $r=\mu+v-1$ .

### C. THE BRANCHING PROCESS IN A RANDOM ENVIRONMENT

Suppose that the environment affects all individuals independently but that it may change randomly itself. Thus at time  $n$  the  $j^{\text{th}}$  individual will be replaced by  $V_{nj}$  individuals, where the  $V_{nj}$  are independent, but their distribution may depend on  $n$ . Here is a simple model with that property.

Let  $Z_i, i=1, 2, \dots$  be positive i.i.d. random variables bounded away from zero, and having finite variance. Let  $Y_{nj}$  be another family of i.i.d. random variables, independent of the  $Z_i$ , with mean zero and variance  $\hat{\sigma}^2$ . Then we set

$$(3.14) \quad N_{n+1} = \sum_{j=1}^N v_{nj} = Z_n N_n + \sum_{j=1}^N Y_{nj}$$

The process  $N_n$  is an example of a branching process in a random environment (BPRE) (Smith and Wilkinson, 1969; Athreya and Karlin, 1971 a,b; Keiding, 1975; Goettge, 1975)\*.

We will show that the diffusion approximation to the BPRE satisfies the Ito equation

$$(3.15) \quad dN = (\mu + 1/2 \sigma^2) N dt + \sqrt{\sigma^2 N^2 + \sigma^2 N} dW \quad (\text{Ito}),$$

where  $\mu = E\{\log Z_1\}$  and  $\sigma^2 = \text{Var}\{\log Z_1\}$ .

It is possible to derive (3.15) directly as an integral equation but this task is considerably more complex than it was in the branching model. Let us rather assume that we know the limit is a diffusion, and simply identify its infinitesimal generator, which means in the present context to identify the drift and diffusion coefficients. This we can do by finding the "infinitesimal mean and

---

\* There are many ways of letting the distribution of  $v_{nj}$  depend on the environment. We have merely chosen one of the simplest.

variance" and appealing to the following result (Gihman and Skorohod, 1972, p. 73). If  $\mu(x)$  and  $\sigma(x)$  are continuous functions such that  $|\mu(x)|^2 + |\sigma(x)|^2 \leq K(1+x^2)$  for some  $K>0$  and all  $x$ , and if  $X_t$  satisfies the Ito equation

$$(3.16) \quad \begin{cases} dX = \mu(X)dt + \sigma(X)dW & \text{(Ito)} \\ X_0 = x_0 \end{cases}$$

then

$$\mu(x_0) = \frac{d}{dt} E\{X_t\} \Big|_{t=0}$$

and

$$\sigma^2(x_0) = \frac{d}{dt} E\{X_t^2\} \Big|_{t=0} = \frac{d}{dt} \text{Var}\{X_t\} \Big|_{t=0}.$$

As we did in the case of the branching process, we can replace (3.14) by

$$(3.17) \quad N_{n+1} = Z_n N_n + \hat{\sigma} \sqrt{N_n} \Delta_n W,$$

where the Brownian motion  $W$  is independent of the  $Z_n$ .

From (3.17) we see that  $E\{N_n | N_{n-1}, \dots, N_0\} = N_{n-1} E\{Z_n\}$ . Since the  $Z_n$  are i.i.d. it follows by induction that

$$(3.18) \quad E\{N_n | N_0\} = N_0 (E\{Z_1\})^n.$$

Furthermore,

$$(3.19) \quad E\{N_n^2 | N_{n-1}, \dots, N_0\} = N_{n-1}^2 E\{Z_{n-1}\} + \hat{\sigma}^2 N_{n-1} E\{(\Delta_{n-1} W)^2\}$$

where we have used the fact that  $Z_n$  and  $\Delta_n W$  are independent. Let  $\alpha = E\{Z_1\}$  and  $\beta = E\{Z_1^2\}$ . By induction, using (3.18)

$$(3.20) \quad E\{N_n^2 | N_0\} = \beta^n N_0^2 + \hat{\sigma}^2 N_0 E\{(\Delta_1 W)^2\} (\alpha^{n-1} + \alpha^{n-2} \beta + \dots + \alpha \beta^{n-1}) \\ = \beta^n N_0^2 + \frac{\beta^n - \alpha^n}{\beta - \alpha} \hat{\sigma}^2 N_0 E\{(\Delta_1 W)^2\}.$$

Now suppose the generations arrive at shorter and shorter intervals, and look at the times  $\epsilon, 2\epsilon, \dots$ . We are led to

$$(3.21) \quad N_{(n+1)\epsilon} = Z_{n\epsilon} N_{n\epsilon} + \hat{\sigma} \sqrt{N_{n\epsilon}} \Delta_{n\epsilon} W,$$

where the  $Z_{n\epsilon}$  are again i.i.d., but their distribution may depend

~~on  $\epsilon$~~  we suppose that as  $\epsilon \rightarrow 0$

$$E\{\log Z_{1\epsilon}\} \sim \epsilon\mu \quad \text{and} \quad \text{Var}\{\log Z_{12}\} \sim \epsilon\sigma^2,$$

where  $\mu$  and  $\sigma^2$  represent the mean and variance of  $\log Z_{1\epsilon}$  "per generation". Setting  $Z_{n\epsilon} = 1 + U_{n\epsilon}$  we will have  $E\{U_{n\epsilon}\} \sim \epsilon(\mu + 1/2 \sigma^2)$  and  $\text{Var}\{U_{n\epsilon}\} \sim \epsilon\sigma^2$ . Let  $\tilde{N}_t$  be the limiting diffusion (which we have assumed to exist). We must compute

$$\nu(\tilde{N}) = \lim_{h \rightarrow 0} \frac{1}{h} E\{\tilde{N}_{t+h} - \tilde{N}_t \mid \tilde{N}_t = \tilde{N}\}$$

Let  $\epsilon > 0$  and  $h > 0$ , and put  $k = h/\epsilon$ . Then

$$h^{-1} E\{\tilde{N}_h - \tilde{N}_0\} = \lim_{\epsilon \rightarrow 0} h^{-1} E\{N_{k\epsilon} - N_0\}$$

By the obvious modification of (3.19) this equals

$$\begin{aligned} &= \lim_{\epsilon \rightarrow 0} h^{-1} N_0 (E\{Z_1\}^k - 1) \\ &= \lim_{\epsilon \rightarrow 0} h^{-1} N_0 ([1 + \epsilon(\mu + 1/2 \sigma^2) + o(\epsilon)]^{h/\epsilon} - 1) \\ &= h^{-1} N_0 (e^{(\mu + 1/2 \sigma^2)h} - 1) \end{aligned}$$

Now let  $h \rightarrow 0$ . This converges to  $K_1(N_0)$  defined by

$$(3.22) \quad K_1(N_0) = (\mu + 1/2 \sigma^2) N_0.$$

Next compute

$$h^{-1} E\{(\tilde{N}_h - \tilde{N}_0)^2\} = h^{-1} \lim_{\epsilon \rightarrow 0} [E\{N_{k\epsilon}^2\} - 2N_0 E\{N_{k\epsilon}\} + N_0^2]$$

Apply (3.19) and (3.20) with  $\alpha = 1 + (\mu + 1/2 \sigma^2) \epsilon + o(\epsilon)$  and  $\beta = 1 + 2\epsilon(\mu + \sigma^2) + o(\epsilon)$ . Then  $\frac{\alpha^k - \beta^k}{\alpha - \beta} = k + o(\epsilon)$ , so

$$\begin{aligned} &= h^{-1} \lim [(1 + 2\epsilon(\mu + \sigma^2))^{h/\epsilon} N_0^2 + \frac{h}{\epsilon} \hat{\sigma}^2 N_{0\epsilon} - 2 N_0^2 (1 + (\mu + 1/2 \sigma^2) \epsilon)^{h/\epsilon} + N_0^2] \\ &= \hat{\sigma}^2 N_0 + h^{-1} (e^{2(\mu + \sigma^2)h} - 2 e^{(\mu + 1/2 \sigma^2)h} - 1) N_0. \end{aligned}$$

As  $h \rightarrow 0$  this converges to a function  $K_2(N_0)$  defined by

$$K_2(N_0) = \hat{\sigma}^2 N_0 + \sigma^2 N_0^2.$$

Thus the limiting diffusion  $\tilde{N}$  satisfies

$$d\tilde{N} = K_1(\tilde{N}) dt + \sqrt{K_2(\tilde{N})} dW,$$

which gives (3.15).

Remarks

1. Keiding (1975) has outlined a derivation similar to the above, and has worked out the boundary classification.

2. The BPRE model shows the relationship between the random environment and the branching models. Both are special cases of the BPRE model. If we suppose that the environment is constant, then  $\sigma^2=0$  and we get the branching model, while if the sampling effects are negligible,  $\hat{\sigma}^2=0$  and we get the random environment model. Notice also that if the population size is large,  $\sqrt{\sigma^2 N^2 + \hat{\sigma}^2 N} \sim \sigma N$ , so  $N$  behaves like the random environment model. On the other hand, for very small  $N$ ,  $\sqrt{\sigma^2 N^2 + \hat{\sigma}^2 N} \sim \hat{\sigma}\sqrt{N}$ , so the BPRE model behaves like the branching model. In particular, the population in the BPRE model can become extinct in finite time, in contrast to the unrealistic behavior of the random environment model, in which the population can tend asymptotically to extinction but never completely dies out.

#### 4. DIFFUSION MODELS FOR GENE FREQUENCIES

The three diffusion approximations we derived above could have been derived from diffusion theory with no reference to a stochastic calculus. However, one of the advantages of stochastic

differential equations is the ease with which they allow one to combine simpler models into more complex ones. We will illustrate this with some examples in population genetics.

Suppose a population consists of  $N_t$  individuals, and consider a haploid gene at a single locus with alleles  $A$  and  $a$ . Let  $L_t$  be the number of  $A$ -genes and let  $M_t$  be the number of  $a$ -genes, so that  $N_t = L_t + M_t$ . We make the following two assumptions:

- (1)  $L_t$  and  $M_t$  each evolve according to one of the models of section 3;
- (2)  $L_t$  and  $M_t$  evolve independently.

One then studies the frequency of  $A$ , which is  $X_t = L_t/N_t$ . Since the BPFE model contains the other two models as special cases, we may as well assume that both  $L_t$  and  $M_t$  evolve according to it.

Now if  $L_t$  and  $M_t$  approximate the true number of  $A$ - and  $a$ -genes, then  $X_t$  will approximate the true gene frequency. To derive the stochastic equation satisfied by  $X_t$ , we will need the two-dimensional version of Ito's formula (McKean, 1969).

Let  $g(x,y)$  be suitably differentiable in both arguments. Then

$$dg(L,M) = g_x(L,M)dL + g_y(L,M)dM + g_{xy}(L,M)(dL)(dM) \\ + 1/2 g_{xx}(L,M)(dL)^2 + 1/2 g_{yy}(L,M)(dM)^2.$$

Applying this to  $X_t = \frac{L_t}{L_t + M_t}$  and noting that

$(dL)(dM) \equiv 0$  by virtue of the independence of  $L$  and  $M$ ,

$$(4.1) \quad dX = \frac{M}{N^2} dL - \frac{L}{N^2} dM - \frac{M}{N^3} (dL)^2 + \frac{L}{N^3} (dM)^2$$

By hypothesis  $L$  and  $M$  satisfy the stochastic differential equations:

$$(4.2) \quad dL = (\mu_1 + 1/2 \sigma_1^2) L dt + \sqrt{\sigma_1^2 L^2 + \hat{\sigma}_1^2} L dW^{(1)} \quad (\text{Ito})$$

$$dM = (\mu_2 + 1/2 \sigma_2^2) M dt + \sqrt{\sigma_2^2 M^2 + \hat{\sigma}_2^2} M dW^{(2)} \quad (\text{Ito})$$

where  $W^{(1)}$  and  $W^{(2)}$  are independent standard Brownian motions and  $\mu_1, \sigma_1$  and  $\hat{\sigma}_1$  are real numbers. From the "multiplication table" of Mc Kean (1969, p.44) we see that  $(dL_t)^2 = \sigma_1^2 L_t^2 + \hat{\sigma}_1^2 L_t$  and

$(dM_t)^2 = \sigma_2^2 M_t^2 + \hat{\sigma}_2^2 M_t^2$  so that on setting  $\sigma^2 = \sigma_1^2 + \sigma_2^2$ ,  $s = \nu_1 - \nu_2$ ,

and utilizing (4.2) we obtain the Ito equation:

$$(4.3) \quad dX = X(1-X) \left[ s + \frac{1}{2} \sigma^2 (1-2X) + \frac{\hat{\sigma}_2^2 - \hat{\sigma}_1^2}{N} \right] dt \\ + (1-X) \sqrt{\frac{\sigma_1^2 X^2 + \hat{\sigma}_1^2 X}{N}} dW^{(1)} - X \sqrt{\frac{\sigma_2^2 (1-X)^2 + \hat{\sigma}_2^2 (1-X)}{N}} dW^{(2)}$$

This equation involves two independent Brownian motions, and we would like to reduce it to an equation involving a single one. This can be done as follows.

Let 
$$f_i(x, N) = (1-x) \sqrt{\frac{\sigma_i^2 x^2 + \hat{\sigma}_i^2 x}{N}}, \quad i = 1, 2.$$

Define a new process  $W_t$  by  $W_0 = 0$  and

$$dW = (f_1^2(x, N) + f_2^2(1-x, N))^{-1/2} [f_1(x, N) dW^{(1)} + f_2(1-x, N) dW^{(2)}].$$

This is again a Brownian motion. (Indeed, a direct calculation shows that  $E\{W_{t+s} - W_t\} = 0$  and  $E\{(W_{t+s} - W_t)^2\} = s$  for all  $s, t \geq 0$ .) In terms of this new Brownian motion we can rewrite

(4.3) as

$$(4.4) \quad dX = X(1-X) \left[ s + \frac{1}{2} \sigma^2 (1-2X) + \frac{\hat{\sigma}_2^2 - \hat{\sigma}_1^2}{N} \right] dt \\ + \sqrt{\sigma^2 X^2 (1-X)^2 + \frac{\hat{\sigma}_2^2 - \hat{\sigma}_1^2}{N} X^2 (1-X) + \frac{\hat{\sigma}_1^2}{N} X(1-X)} dW \quad (\text{Ito}),$$

where  $s = \mu_1 - \mu_2$  and  $\sigma^2 = \sigma_1^2 + \sigma_2^2$ . This is a rather complicated formula, but it simplifies greatly in many special cases. Note that  $\hat{\sigma}_1$  and  $\hat{\sigma}_2$  tend to cancel each other out. If  $\hat{\sigma}_1 = \hat{\sigma}_2 = \hat{\sigma}$ , say, then

$$(4.5) \quad dX = X(1-X) \left[ s + \frac{1}{2} \sigma^2 (1-2X) \right] dt + \sqrt{\hat{\sigma}^2 X^2 (1-X)^2 + \frac{\hat{\sigma}^2}{N} X(1-X)} dW$$

If both L and M evolve according to the random environment model, then  $\hat{\sigma}_i^2 = 0$ , and

$$(4.6) \quad dX = X(1-X) \left[ s + \frac{1}{2} \sigma^2 (1-2X) \right] + \sigma X(1-X) dW \quad (\text{Ito}),$$

which is equivalent to the Stratonovich equation

$$(4.7) \quad dX = sX(1-X) + \sigma X(1-X) dW \quad (\text{Stratonovich})$$

These equations (4.6) and (4.7) agree with those derived directly from a discrete model by Gillespie (1973) and studied by Tuckwell (1976). These authors showed that "quasi

fixation" occurs for A if  $\mu_1 > \mu_2$ . They differ in their drift term from those studied by Kimura (1954, 1964) and Wright (1948). (We point out that Kimura and Wright studied  $X_t$  via diffusion process theory rather than stochastic differential equations. It is perhaps misleading to compare our results with theirs, for their equations are derived under the assumption that  $N$  is constant, whereas in our case - and Gillespie's -  $N$  is not constant.)

Another interesting specialization occurs when  $L$  and  $M$  evolve according to the branching model. In this case  $\sigma^2 = 0$  so

$$(4.8) \quad dX = \left( s + \frac{\hat{\sigma}_1^2 - \hat{\sigma}_2^2}{N} \right) X(1-X) dt + \sqrt{\frac{\hat{\sigma}_1^2 X(1-X)^2 + \hat{\sigma}_2^2 X^2(1-X)}{N}} dW \quad (\text{Ito})$$

It is biologically reasonable to assume that

$\hat{\sigma}_1 = \hat{\sigma}_2 = \hat{\sigma}$  in which case (4.8) becomes

$$(4.9) \quad dX = sX(1-X) dt + \sigma \sqrt{\frac{X(1-X)}{N}} dW \quad (\text{Ito}),$$

or, in the Stratonovich calculus:

$$(4.10) \quad dX = \left( sX(1-X) + \frac{1-2X}{2N} \right) dt + \sigma \sqrt{\frac{X(1-X)}{N}} dW. \quad (\text{Stratonovich})$$

We should point out that while equation (4.4) contains many other special cases, it may arise in its unsimplified form in many biologically interesting instances. For example, if one of the alleles confers resistance to a hostile environment - adaptation to extreme cold would be a case in point - we might have  $\sigma_1 \ll \sigma_2$ .

A few remarks are in order here. These equations are formally the same as those derived by Kimura (1955, 1962, 1964), Wright (1948, 1964) and more recently by Ohta (1972), in which the sampling is assumed to be binomial so that the rate of change of variance of  $X$  is  $\frac{1}{N} X(1-X)$ . There is a rather puzzling aspect of this, however. Their models were derived under the assumption that  $N_t$  is constant, whereas  $N$  varies in our model. Is this coincidence or not? Perhaps there is a way to answer this. It is possible to compute the conditional distribution of  $L_t$  given that  $N_t$  is constant; it turns out that this conditional process satisfies a certain stochastic differential equation, and one can then compute the stochastic differential equation satisfied by  $X = L/N$ . It is identical to (4.9)! In other words the behavior of  $X$ , conditioned on  $dN=0$  is given by (4.9). (This doesn't quite answer the question of whether or not this is due to coincidence: if one does the same thing to either the random environment or BPRE model, one ends up with a nearly unrelated stochastic differential

equation for  $X$ , so that the conditional and unconditional behavior of  $X$  are governed by the same equation only in the branching model).

A second surprising feature of the above models is that, whereas  $X_t$  is a diffusion process in the random environment model (cf(4.6)) it is not -and in fact it is not even a Markov process- in either the branching or BPRE model, cf (4.5) and (4.9). This is because  $N_t$ , which is not constant, appears in both equations. The pair  $(L_t, N_t)$  is, however, a Markov process, and it is this pair which should be the basic object of study. One can still learn about the behavior of  $X$  from its differential equation. For example, in the branching model (4.5), if the population size is large the diffusion coefficient is small and the behavior of  $X$  is nearly deterministic, so we have

$$(4.11) \quad \frac{dX}{dt} \sim sX_t(1-X_t),$$

and we expect to have either loss or fixation of the allele  $A$  according to whether  $s$  is less or greater than zero.

Certain other facts are more easily seen by regarding the system (4.2). For instance—and this holds in either

the branching or BPRE model-- there are two ways for  $X$  to tend to either zero or one. First, either  $A$  or  $a$  may become extinct. If this happens, it happens in finite time, and hence  $X$  will reach either 0 or 1 in finite time. On the other hand, both  $L_t$  and  $M_t$  may tend to infinity, in which case  $X$  may converge to either 0 or 1 as  $t \rightarrow \infty$ , without ever reaching its limit.

We note that the special case of (4.4) obtained by setting  $s=0$  and  $\hat{\sigma}_1 = \hat{\sigma}_2$  has been derived by a different method by Jensen and Pollack (1969) and Jensen (1973). These authors compared the solutions of the corresponding Kolmogorov forward equations with a computer calculation on the equivalent Markov chain and also determined the probability of ultimate fixation of an allele.

Finally, it is interesting to note that the model for gene frequency, in which each allele multiplies according to a branching process in a random environment, is an additive version of the branching and the random environment models in the sense that the first and second infinitesimal moments of  $X_t$  in the third model are the sums of the corresponding quantities in the first two models. This was also true for the models of Malthusian growth, but it was not obvious that it would hold for gene frequency too. This supports the procedures of earlier workers (Kimura, 1970 and references therein) in

finding the drift and diffusion coefficients of the gene frequency when there are several sources of evolutionary pressure and several sources of variability. We should point out, however, that care should be exercised when adopting that procedure: additivity may not apply when, for example, mutation from L to M and back is included in the original equations for  $L_t$  and  $M_t$ .

### 5. Discussion

The stochastic equation we derived in section 3 for the BPRE model of unrestricted population growth can be written

$$(5.1) \quad N_t = N_0 + \int_0^t (\mu + 1/2 \sigma^2) N_s ds + \int_0^t \sqrt{\sigma^2 N_s^2 + \hat{\sigma}^2 N_s} dW_s \quad (\text{Ito})$$

If we take expectations on both sides and use the fact that the expectation of an Ito integral is zero we see that

$$E\{N_t\} = N_0 + \int_0^t (\mu + 1/2 \sigma^2) E\{N_s\} ds,$$

from which it follows that

$$(5.2) \quad E\{N_t\} = N_0 e^{(\mu + 1/2 \sigma^2)t} = N_0 e^{at},$$

where  $\alpha = \mu + 1/2 \sigma^2$ . Since the BPRE model includes the random environment and the branching models as special cases, (with  $\sigma = 0$  in the branching model), it is apparent that the expected population size is the same in all three models. If we regard the expected population size as defining a "deterministic model" we can say that all three of our stochastic models correspond to the single deterministic model  $\dot{N} = \alpha N$ , or, in differential form, to

$$(5.3) \quad dN = \alpha N dt.$$

This is, of course, the classical model of Malthusian growth.

This model can be misleading. For instance, in all three models the population will die out with probability one if  $\mu < 0$  and will tend to infinity with positive probability if  $\mu > 0$ . (This is evident in the branching model, for instance, from equation (3.3) ) On the other hand the expected population size will tend to either zero or infinity according to whether  $\mu + 1/2\sigma^2$  is negative or positive respectively. It is certainly possible to have  $\mu < 0$  and  $\mu + 1/2 \sigma^2 > 0$ , in which case the expected population size goes to infinity while the population itself dies out. This is no paradox. What is happening is that, even though  $N_t$  may equal zero with high probability, if it is not zero, it is quite likely to be

extremely large, and this probability is great enough to make its expectation tend to infinity.

Suppose we had started studying population size with the classical equation (5.3) and had formed a stochastic model by "throwing in white noise". We would have simply replaced  $adt$  by  $\sigma dW + \mu dt$  in (5.3), and would have gotten the random environment model as a result. Apart from the fact that we would then be faced with a choice of calculi, we would have lost the richness of the branching and the BPRE models. This would be a real loss, for the second two models have properties quite different from those of the first. For example,  $N = 0$  is always accessible in these two models whereas it is inaccessible in the random environment model - a biologically unrealistic behavior since we know that a real population can become extinct.

This all suggests that the practice of obtaining a stochastic model by randomizing an existing deterministic model is naive. Rather than arbitrarily putting noise in a deterministic system, one should lavish at least as much care in analyzing how the randomness enters as in determining how the system behaves without noise.

We have derived several diffusions as limits of discrete parameter processes. In the usual situation, however, one has a fixed process, in either discrete or continuous time, and one wishes to approximate it by continuous process.

No limit is involved; the limit theorem is merely used to justify the approximation if, say, the population size is large. Such approximations have been found useful and quite accurate in population genetics, even when the population size is small and one would think that the limiting argument doesn't hold (Watterson, 1962; Karlin & McGregor, 1964; Ewens, 1965).

The approximations we have considered are of discrete processes, and the reader will have remarked that the stochastic differential equations which resulted were Ito equations rather than Stratonovich equations. This was to be expected. Suppose, for instance, that for each  $\epsilon > 0$  one has a process satisfying the difference equation

$$N_{(n+1)\epsilon} - N_{n\epsilon} = f(N_{n\epsilon})\epsilon + g(N_{n\epsilon}) \Delta_{n\epsilon} W.$$

If  $N_{n\epsilon}$  converges to a diffusion as  $\epsilon$  approaches zero, then, subject to some technical conditions on  $f$  and  $g$ , this diffusion will

satisfy the Ito equation

$$(5.4) \quad dN = f(N)dt + g(N)dW \quad (\text{Ito}).$$

A quite different situation arises when we have a process  $N(t)$  in continuous time to approximate. Suppose it satisfies

$$(5.5) \quad \frac{dN(t)}{dt} = f(N(t)) + g(N(t)) \frac{d\hat{W}(t)}{dt}$$

where  $\hat{W}$  is a process smooth enough so that its derivative exists almost everywhere. We can think of  $\frac{d\hat{W}}{dt}$  as the "noise" in the system.

It is evidently far from white - and  $\hat{W}$  is not a Brownian motion (for Brownian motion is not differentiable). However, by the convergence theorem of Wong and Zakai (1965), if  $\hat{W}(t)$  approximates a Brownian motion in distribution, then  $N(t)$  will be approximated by the solution  $\tilde{N}$  of the Stratonovich equation

$$(5.6) \quad d\tilde{N} = f(\tilde{N}) dt + g(\tilde{N}) dW \quad (\text{Stratonovich})$$

Note that this is a totally different type of approximation from the first. In the first case we approximated the solution of a difference equation by a diffusion, and we ended up with an Ito equation. In the second case we approximated a smooth process

$\hat{W}_t$  by a very unsmooth process, Brownian motion; in other words we approximated correlated noise by white noise, and the result was a Stratonovich equation.

To conclude, it is apparent that both Ito and Stratonovich integrals can arise naturally. The continued controversy over the correct "interpretation" of certain stochastic differential equations is not evidence that stochastic integrals are pathological, and even less that one calculus is more suitable to biology than the other; it is merely evidence that the models themselves have been derived with too little care.

References

- Athreya, K.B. and Karlin, S. 1971a. On branching processes with random environments I: Extinction Probabilities. *Ann. Math. Statist.* 42, 1499-1520.
- Athreya, K.B. and Karlin, S. 1971b. On branching processes with random environments II: Limit Theorems. *Ann. Math. Statist.* 42, 1843-1858.
- Cook, R.D. and Hartl, D.L. 1974. Uncorrelated random environments and their effects on gene frequency. *Evolution* 28, 265-74.
- Crow, J.F. and Kimura, M. 1970. "An Introduction to Population Genetics Theory". Harper and Row, New York.
- Doob, J.L. 1953. "Stochastic Processes". Wiley, New York.
- Ewens, W.J. 1965. The adequacy of the diffusion approximation to certain distributions in genetics. *Biometrics* 21, 386-394.
- Feldman, M.W. and Roughgarden, J. 1975. A population's stationary distribution and chance of extinction in stochastic environment with remarks on the theory of species packing. *Theor. Pop. Biol.* 7, 197-207.
- Feller, W. 1951. Diffusion processes in genetics, in *Proc. Second Berkeley Symposium Math. Stat. Prob.*, 227-246. University of California Press, Berkeley.
- Gihman, I.I. and Skorohod, A.V. 1972. "Stochastic Differential Equations". Springer-Verlag, New York, Heidelberg, Berlin.
- Gillespie, J.H. 1972. The effects of stochastic environments on allele frequencies in natural populations. *Theor. Pop. Biol.* 3, 241-248.
- Gillespie, J.H. 1973. Natural selection with varying selection coefficients - a haploid model. *Genet. Res.* 21, 115-120.
- Goel, N.C., Maitra, S.C. and Montroll, E.W. 1971. On the Volterra and other non-linear models of interacting populations. *Rev. Mod. Phys.* 43, 231-276.

- Goettge, R.T. 1975. Limit theorems for the supercritical Galton-Watson process in varying environments. *Math. Biosci.* 28, 171-190.
- Gray, A.H. and Caughey, T.K. 1965. A controversy in problems involving random parametric excitation. *J. Math. and Phys.* 44, 288-296.
- Ito, K. 1951a. On Stochastic Differential Equations. *Mem. Am. Math. Soc.* No. 4.
- Ito, K. 1951b. On a formula concerning stochastic differentials. *Nagoya Math. J.* 3, 55-65.
- Jaswinski, A.H. 1970. "Stochastic processes and Filtering Theory". Academic Press, New York.
- Jensen, L. 1973. Random selective advantages of genes and their probabilities of fixation. *Genet. Res.* 21, 215-219.
- Jensen, L. and Pollack, E. 1969. Random selective advantages of a gene in a finite population. *J. Appl. Prob.* 6, 19-37.
- Karlin, S. and McGregor, J. 1964. On some stochastic models in genetics, in, "Stochastic Models in Medicine and Biology", (J. Gurland, Ed.). University of Wisconsin Press, Madison.
- Keiding, N. 1975. Extinction and exponential growth in random environments, *Theor. Pop. Biol.* 8, 49-63.
- Kiester, A.R. and Barakat, R. 1974. Exact solutions to certain stochastic differential equation models of population growth, *Theor. Pop. Biol.* 6, 199-216.
- Kimura, M. 1954. Process leading to quasifixation of genes in natural populations due to random fluctuations of selection intensities. *Genetics* 39, 280-294.
- Kimura, M. 1955a. Process of random genetic drift with a continuous model. *Proc. Nat. Acad. Sci. Wash.* 42, 336-340.
- Kimura, M. 1955b. Stochastic processes and distribution of gene frequencies under natural selection. *Cold Spring Harbor Symp. Quant. Biol.* 20, 33-53.
- Kimura, M. 1962. On the probability of fixation of mutant genes in a population. *Genetics* 47, 713-719 (1962).

- Kimura, M. 1964. Diffusion models in population genetics. *J. Appl. Prob.* 1, 177-232.
- Kimura, M. 1970. Stochastic processes in population genetics, with special reference to distribution of gene frequencies and probability of gene fixation, *in* *Biomathematics Volume 1, Mathematical Topics in Population Genetics*, (K. Kojima, Ed.), Springer-Verlag, New York, Heidelberg, Berlin.
- Levins, R. 1969. The effects of random variations of different types on population growth. *Proc. Nat. Acad. Sci.* 62, 1061-1065.
- Lewontin, R.C. and Cohen, D. 1969. On population growth in randomly varying environment. *Proc. Nat. Acad. Sci.* 62, 1056-1060.
- Ludwig, D. 1976. A singular perturbation problem in the theory of population extinction, in *AMS conference on Asymptotic Methods and Singular Perturbations*, New York (in press).
- May, R.M. 1971. Stability in model ecosystems. *Procl. Ecol. Soc. Aust.* 6, 18-56.
- May, R.M. 1973. "Stability and Complexity in Model Ecosystems", Princeton University Press, Princeton, New Jersey.
- McKean, H.P. 1969 "Stochastic Integrals", Academic Press, New York.
- Mortensen, R.E. 1969. Mathematical Problems of modeling stochastic nonlinear dynamic systems, *J. Stat. Phys.* 1, 271-296.
- Ohta, T. 1972. Fixation probability of a mutant influenced by random fluctuation of selection intensity, *Genet. Res.* 19, 33-38.
- Smith, W.L. and Wilkinson, W.E. 1969. On branching processes in random environments, *Ann. Math. Statist.* 40, 814-827.
- Stratonovich, R.L. 1966. A new representation for stochastic integrals and equations, *SIAM J. Control* 4, 362-371.
- Tuckwell, H.C. 1974. A study of some diffusion models of population growth, *Theor. Pop. Biol.* 5, 345-357.
- Tuckwell, H.C. 1976. *The effects of random selection on gene frequency*; *Math. Biosci.* (in press).

- Watterson, G.A. 1962. Some theoretical aspects of diffusion theory in population genetics, *Ann. Math. Statist.* 33, 939-957.
- Wong, E. and Zakai, M. 1965. On the convergence of ordinary integrals to stochastic integrals, *Ann. Math. Statist.* 36, 1560-1564.
- Wright, S. 1948. On the roles of directed and random changes in the genetics of populations, *Evolution* 2, 279-295.
- Wright, S. 1964. Stochastic processes in evolution in "Stochastic model in Medicine and Biology", (J. Gurland, Ed.), University of Wisconsin Press, Madison.