The Annals of Applied Probability
1998, Vol. 8, No. 1, 182–192

ELLIPTIC AND OTHER FUNCTIONS IN THE LARGE DEVIATIONS BEHAVIOR OF THE WRIGHT–FISHER PROCESS

BY F. PAPANGELOU

University of Manchester

The present paper continues the work of two previous papers on the variational behavior, over a large number of generations, of a Wright–Fisher process modelling an even larger reproducing population. It was shown that a Wright–Fisher process subject to random drift and one-way mutation which undergoes a large deviation follows with near certainty a path which can be a trigonometric, exponential, hyperbolic or parabolic function. Here it is shown that a process subject to random drift and gamete selection follows in similar circumstances a path which is, apart from critical cases, a Jacobian elliptic function.

1. Introduction. This is the third in a series of papers dealing with the variational problem associated with the large deviations behavior of a Wright–Fisher process modelling the genetic evolution of a finite population. The problem in question, familiar from the Freidlin–Wentzell large deviations theory, is that of determining the "preferred paths" of the process, that is, paths followed with near certainty by the suitably scaled process when its state undergoes a large deviation. The first paper [5] dealt with this problem in the case of a process subject only to random drift (arising from random sampling), where the preferred paths were shown to be trigonometric cosines. The second paper [6] treated the case where one-way mutation is added to the random drift and revealed a richness and diversity of preferred paths out of all proportion to the simplicity of the process involved; a preferred path can be a trigonometric, exponential, hyperbolic or parabolic function, depending on the value of the mutation parameter and the boundary conditions. With the present paper the picture becomes even more intricate; as we will show, a fascinating pattern of mostly elliptic (Jacobian) functions appears when the process is subject to random drift and gamete selection.

In the context of the classical Freidlin–Wentzell theory, a randomly perturbed dynamical system undergoing a large deviation follows with near certainty a path minimizing the action functional. For the Wright–Fisher process it is not the large deviations theory for diffusions which is relevant but rather its variant for discrete time Markov processes developed by Wentzell (see [7]). It was shown in [5] and [6] that the variational aspects of this theory can be adapted to apply to the Wright–Fisher process if the latter is scaled in such a manner that the number of generations per "unit of time," though large, is

Received March 1997; revised June 1997.

AMS 1991 *subject classifications*. Primary 60F10; secondary 60J20.

Key words and phrases. Wright–Fisher process, large deviations, action functional, calculus of variations, elliptic functions.

very much smaller than the size of the population and any forces of mutation or selection act on the generational time scale and are thus stronger than genetic drift (sampling noise). This feature of stronger-than-drift mutation and selection is shared with [4] (cf. [5]), where, however, the authors used large deviations arguments to tackle a different problem: the asymptotics of the exit time of a generalization of the Wright–Fisher process from a neighborhood of a stable equilibrium point. (An extension of this to an infinite-alleles analogue can be found in [3]).

It is worth mentioning that the Wright–Fisher process with selection was used (but not in a large deviations context) by Kaplan and Darden in [2] to illustrate a general result of theirs on the manner in which a small-noise Markov chain follows its deterministic approximation in the neighborhood of an equilibrium.

More details of the variational problem treated in the present paper are given in the next section where we state the basic definitions and technical facts needed and review briefly the preferred paths uncovered in [5] and [6] for the cases referred to above. This will provide the right backdrop for a good understanding of the pattern of paths that will emerge later in the case of selection. It transpires that, unlike the case of mutation, there can be more than one extremal path (i.e., solution of the relevant Euler equation) joining two points and staying clear of the boundary at 0 and 1. It is then necessary to determine which of these extremals actually minimize the pseudo-action-functional, a task taken up in Section 4. The paper concludes with the main result on the concentration of probability, stated as Theorem 4.1.

2. Preliminaries. The main result of [6], stated as Theorem 2.1 below, is a limit theorem for a sequence $Y_t^{(n)}$, $t \geq 0$; $n = 1, 2, \ldots$ of increasingly severe scalings of the Wright–Fisher process, which are introduced as continuous time processes as follows. For each $n \geq 1$, the process $Y_t^{(n)}$, $t \geq 0$ jumps at times $1/n, 2/n, 3/n, \ldots$ and is constant on any of the intervals $[k/n, (k+1)/n)$, $k = 0, 1, 2, \ldots$. The skeleton process $Y_0^{(n)}, Y_{1/n}^{(n)}, Y_{2/n}^{(n)}, \ldots$ is a Wright–Fisher process, with $Y_{k/n}^{(n)}$ representing the proportion of A-alleles, say, in the $(k+1)$th generation of a population of $2N$ genes, where $N = N(n)$ depends on n. *It will be assumed throughout that $N/n \to \infty$ as $n \to \infty$.* The one-step transition probability $P(y, \tilde{y})$ of the skeleton process is, with $\tilde{y} = j/2N$, given by

$$P(y, \tilde{y}) = \binom{2N}{j} p_{y,n}^j (1 - p_{y,n})^{2N-j}$$

where $p_{y,n}$ is in general of the form

$$p_{y,n} = y + \frac{g(y)}{n} + o\left(\frac{1}{n}\right) \tag{2.1}$$

with $0 \leq p_{y,n} \leq 1$, $g(y)$ a continuous function on $[0, 1]$ and $o(1/n)/(1/n) \to 0$ uniformly in $y \in [0, 1]$. The case of pure random drift discussed in [5] corresponds to $p_{y,n} = y$, while that of one-way mutation treated in [6] corresponds to $p_{y,n} = y(1-\gamma/n)$ where $\gamma > 0$. Two-way mutation has $p_{y,n} = y + n^{-1}[(1-y)\gamma_2 - y\gamma_1]$ ($\gamma_1 > 0$, $\gamma_2 > 0$). If there is no mutation but there are selective forces working against the A-allele, then $p_{y,n}$ can be taken to be

$$p_{y,n} = y\left[y + \left(1+\frac{\beta}{n}\right)(1-y)\right]^{-1} = y - \frac{\beta}{n}y(1-y) + o\left(\frac{1}{n}\right), \tag{2.2}$$

where $\beta > 0$. This is the case we will investigate below. The more general case of zygotic selection in a diploid population involves a cubic $g(y)$ and will not be dealt with here.

For given $y \in [0,1]$, let $G(y,z)$ be the cumulant generating function of the Gaussian distribution with mean $g(y)$ and variance $y(1-y)$, that is, $G(y,z) = g(y)z + \frac{1}{2}y(1-y)z^2$ and consider also the cumulant of $Y^{(n)}_\bullet$, that is, the function $G^n(y,z) := n\log E_y \exp\{z(Y^{(n)}_{1/n} - y)\}$, which can easily be calculated explicitly. It is shown in [6] that the function $n(2N)^{-1}G^n(y, 2Nn^{-1}z)$ does not depend on N, that this function and its derivative with respect to z converge, as $n \to \infty$, to $G(y,z)$ and its derivative, respectively, and that certain uniformity and boundedness conditions required in Wentzell's theory [7] are satisfied. A large deviations principle cannot however be deduced for $Y^{(n)}_\bullet$ because the Legendre transform of $G(y,z)$

$$H(y,u) = \sup_z[zu - G(y,z)] = \frac{1}{2}\frac{(u-g(y))^2}{y(1-y)} \tag{2.3}$$

has singularities at $y = 0$ and $y = 1$. Such a principle can be deduced for a modified sequence $\tilde{Y}^{(n)}_\bullet$ of processes, where $\tilde{Y}^{(n)}_t$ behaves as $Y^{(n)}_t$ as long as it is in the range $[\varepsilon, 1-\varepsilon]$ for suitably small $\varepsilon > 0$, but its jumps from a state $y < \varepsilon$ or $y > 1-\varepsilon$ have a Gaussian distribution with mean $g(\varepsilon)/n$ or $g(1-\varepsilon)/n$, respectively, and variance $\varepsilon(1-\varepsilon)/2N$. The modified sequence satisfies on any bounded interval $[0,T]$ a uniform large deviations principle with action functional $(2N/n)\tilde{S}_{0,T}(\phi)$, where $\tilde{S}_{0,T}(\cdot)$ is one of two functionals to be defined presently.

If ϕ is an absolutely continuous function on $[0,T]$, we set

$$S_{0,T}(\phi) = \int_0^T H(\phi(t), \phi'(t))\,dt = \int_0^T \frac{1}{2}\frac{(\phi'(t) - g(\phi(t)))^2}{\phi(t)(1-\phi(t))}\,dt \tag{2.4}$$

(provided $0 \le \phi \le 1$) and

$$\tilde{S}_{0,T}(\phi) = \int_0^T \tilde{H}(\phi(t), \phi'(t))\,dt, \tag{2.5}$$

where $\tilde{H}(y,u)$ agrees with $H(y,u)$ in (2.3) if $\varepsilon \le y \le 1-\varepsilon$ but is equal to $H(\varepsilon,u)$ or $H(1-\varepsilon,u)$ if $y < \varepsilon$ or $y > 1-\varepsilon$, respectively. If ϕ is not absolutely continuous on $[0,T]$, then $S_{0,T}(\phi) = \tilde{S}_{0,T}(\phi) = \infty$.

For more details and a precise statement of some of the implications of the above, the reader is referred to [6]. Here it is sufficient to repeat the informal statement that if $\phi(t)$, $0 \le t \le T$ is an absolutely continuous function which satisfies $\phi(0) = y_0 = Y_0^{(n)}$ and stays clear of 0 and 1, then the logarithm of the probability that $Y_t^{(n)},\ 0 \le t \le T$ follows closely the path $\phi(t),\ 0 \le t \le T$ is of order

$$-Nn^{-1}\int_0^T \frac{(\phi'(t) - g(\phi(t))^2}{\phi(t)(1-\phi(t))}\,dt.$$

See Proposition 2.1 of [6] for the exact statement.

The variational problem, that is, the problem of determining the functions ϕ which minimize $S_{0,T}(\phi)$, was solved for $g(y) \equiv 0$ and $g(y) = -\gamma y$ $(\gamma > 0)$ in [5] and [6], respectively, where it was also proved that the minimizing functions are indeed the "preferred paths" in the probabilistic sense. We summarize here the types of preferred paths ϕ leading from a point $(0, y_0)$ to other points (T, y_1) $(0 < y_1 < 1)$. The description is given in terms of $\phi(0) = y_0$ and $\phi'(0)$ rather than $\phi(0)$ and $\phi(T)$ and the paths should be understood as terminated where they exit from $\Delta = \{(t, y)\colon t \ge 0,\ 0 < y < 1\}$. A fuller description can be found in [5] and [6].

If $g(y) \equiv 0$, then the preferred path ϕ is a monotone arc of the form $\phi(t) = \frac{1}{2} - \frac{1}{2}\cos(ct - \tilde{c})$. If $g(y) = -\gamma y$, there are six different classes of preferred paths, depending on the values of γ and $\phi'(0)$. (i) If $\phi'(0) = 0$ then $\phi(t) = \frac{1}{2} y_0[\cosh \gamma t/\sqrt{1-y_0} + 1]$; (ii) if $0 < |\phi'(0)| < \gamma y_0$ then $\phi(t)$ is of the form

$$\frac{-c}{2(\gamma^2 - c)}\Big[\cosh\Big(\Big(\sqrt{\gamma^2 - c}\,\Big)(t - \tilde{c})\Big) + 1\Big];$$

(iii) if $|\phi'(0)| = \gamma y_0$ then $\phi(t) = y_0 \exp(\pm\gamma t)$; (iv) if $\gamma y_0 < |\phi'(0)| < \gamma\sqrt{y_0}$ then $\phi(t)$ is of the form

$$\frac{c}{2(\gamma^2 - c)}\Big[\cosh\Big(\Big(\sqrt{\gamma^2 - c}\,\Big)(t - \tilde{c})\Big) - 1\Big];$$

(v) if $|\phi'(0)| = \gamma\sqrt{y_0}$ then $\phi(t) = \frac{1}{4}(\gamma t \pm 2\sqrt{y_0})^2$; (vi) if $|\phi'(0)| > \gamma\sqrt{y_0}$ then ϕ is a monotone arc of

$$\frac{c}{2(c - \gamma^2)}\Big[1 - \cos\Big(\Big(\sqrt{c - \gamma^2}\,\Big)(t - \tilde{c})\Big)\Big].$$

THEOREM 2.1 [6]. *If $0 < y_0 < 1$, $T > 0$ and $y_0 e^{-\gamma T} \le y_1 < 1$, then there is a unique ϕ_0 from among the functions described in* (i)–(vi) *above, such that $\phi_0(0) = y_0$ and $\phi_0(T) = y_1$. Assume $Y_0^{(n)} = y_0$ for all n and denote by P the probability measure arising from these initial conditions and the transition structure of $Y_t^{(n)},\ t \ge 0$ corresponding to the case of one-way mutation $g(y) = -\gamma y$ $(\gamma > 0)$. Then, for every $\delta > 0$,*

$$\lim_{n\to\infty} P\Big(\sup_{0 \le t \le T} |Y_t^{(n)} - \phi_0(t)| < \delta \mid Y_T^{(n)} \ge y_1\Big) = 1.$$

The purpose of the present paper is to prove an analogous theorem for the case of selection. Before concluding this section, we formulate a lemma that will be needed below.

LEMMA 2.2. *Suppose that* $g(y) = -\beta y(1-y)$ $(\beta > 0)$ *and that* $0 < y_0 < 1$, $0 < y_1 < 1$, $T > 0$. *If the functional* $S_{0,T}(\phi)$, *restricted to the set of functions* ϕ *on* $[0, T]$ *such that* $\phi(0) = y_0$, $\phi(T) = y_1$ *and* $0 \le \phi \le 1$, *has a minimum at* ϕ_0, *then* $0 < \phi_0(t) < 1$ *for all* $t \in [0, T]$ *and* ϕ_0 *is continuously differentiable on* $[0, T]$ *and therefore satisfies Euler's equation* $H_y - (d/dt)H_{y'} = 0$ *for* $H(y, y')$ *at every* $t \in [0, T]$.

To see why ϕ_0 stays clear of 0 and 1, note that functions of the form $\phi(t) = \tilde{c}(\tilde{c} + \exp(\beta t))^{-1}$ are zero-action functions in the sense of satisfying $H(\phi(t), \phi'(t)) = 0$ at every point. If ϕ_1, ϕ_2 are two such functions satisfying $\phi_1(0) < y_0$, $\phi_1(T) < y_1$ and $\phi_2(0) > y_0$, $\phi_2(T) > y_1$, then $\phi_1(t) \le \phi_0(t) \le \phi_2(t)$ for all $t \in [0, T]$, since otherwise we would be able to reduce the value of $S_{0,T}(\phi_0)$ by replacing a segment of ϕ_0 by a segment of ϕ_1 or ϕ_2. The continuous differentiability of ϕ_0 can be proved by arguments similar to those given in [6] for the case $g(y) = -\gamma y$. See Lemma 3.4 there and the entire paragraph that follows it.

Note that if $\varepsilon > 0$ is chosen so that $\varepsilon \le \phi_1(t) \le \phi_2(t) \le 1-\varepsilon$ for all $t \in [0, T]$ then, subject to $\phi(0) = y_0$, $\phi(T) = y_1$, $\tilde{S}_{0,T}(\cdot)$ too has a minimum at ϕ_0.

3. The extremals. From now on we deal exclusively with a Wright–Fisher process involving selection unfavorable to the A-allele, but no mutation. In this case $g(y) = -\beta y(1-y)$ and hence

$$H(y, u) = \frac{1}{2}\frac{(u + \beta y(1-y))^2}{y(1-y)}$$

and

$$S_{0,T}(\phi) = \frac{1}{2}\int_0^T \frac{(\phi'(t) + \beta\phi(t)(1-\phi(t)))^2}{\phi(t)(1-\phi(t))}\,dt. \tag{3.1}$$

The extremals, that is, the solutions of the corresponding Euler equation, can be obtained from the differential equation

$$\frac{y'^2}{y(1-y)} - \beta^2 y(1-y) = c, \tag{3.2}$$

which expresses the constancy of the "energy" $y'H_{y'} - H$ along each extremal. Although every nonsingular solution of such an equation is an extremal, this is not in general true of singular solutions. Of all constant functions, only the function $y(t) \equiv \frac{1}{2}$, corresponding to $c = -\beta^2/4$, is an extremal, as can be checked directly. (The Euler equation is too long to write here explicitly.)

Let us first describe the solutions of (3.2) without reference to initial values at $t = 0$. Clearly the smallest possible value of c is $-\beta^2/4$ and for this value $y(t) \equiv \frac{1}{2}$ is the only solution since $-\beta^2 y(1-y) \le -\beta^2/4$ implies $y(1-y) = \frac{1}{4}$. Values of c greater than $-\beta^2/4$ will be split into three cases: (I) $c > 0$, (II) $c = 0$ and (III) $-(\beta^2/4) < c < 0$ and be discussed separately. However, first note that

(3.2) can be written in the form

$$y'^2 = \beta^2 y(y-1)\left(y - \frac{1}{2} + \frac{\lambda}{2}\right)\left(y - \frac{1}{2} - \frac{\lambda}{2}\right),$$

where $\lambda = (1 + 4c\beta^{-2})^{1/2}$. Setting $k = \lambda^{-1}$ and using the transformation $z = 2(y - \frac{1}{2})$, we arrive at

$$z'^2 = \frac{\beta^2}{4k^2}(1-z^2)(1-k^2z^2). \tag{3.3}$$

(I) If $c > 0$, then $0 < k < 1$ and if a function $z(t)$ satisfies (3.3) then $w(t) = z(2k\beta^{-1}t)$ satisfies the differential equation

$$w'^2 = (1-w^2)(1-k^2w^2).$$

This shows that

$$w(t) = \operatorname{sn}(t - \tilde{c}, k)$$

where $\tilde{c}$ is a constant and $x = \operatorname{sn}(u, k)$ denotes the periodic Jacobian elliptic function of u which inverts the function

$$u = \int_0^x ((1-s^2)(1-k^2s^2))^{-1/2}\, ds$$

of $x \in [0, 1]$ and is extended to the whole real line so as to be an odd periodic function in the manner of $\sin u$; see [1]. Indeed, the reader who is not familiar with $\operatorname{sn}(\cdot)$ may simply visualize it as a function akin to a sine function, oscillating between -1 and 1. The period of $\operatorname{sn}(\cdot, k)$ is $4K^*$, where

$$K^* = \int_0^1 ((1-s^2)(1-k^2s^2))^{-1/2}\, ds.$$

Returning to (3.2), we see that its solutions are of the form

$$y(t) = \frac{1}{2} + \frac{1}{2}\operatorname{sn}\left(\frac{\beta}{2k}(t - \tilde{c}), k\right) \tag{3.4}$$

with period

$$8k\beta^{-1}\int_0^1 ((1-s^2)(1-k^2s^2))^{-1/2}\, ds. \tag{3.5}$$

As k increases from 0 to 1 this period increases from 0 to ∞.

(II) If $c = 0$, then (3.2) becomes

$$y'^2 = \beta^2 y^2(1-y)^2.$$

The solutions of $y' = \beta y(1-y)$ are of the form $y(t) = \tilde{c}\,(\tilde{c} + e^{-\beta t})^{-1}$ $(\tilde{c} > 0)$, while the solutions of $y' = -\beta y(1-y)$ are of the form $y(t) = \tilde{c}(\tilde{c}+e^{\beta t})^{-1}$ $(\tilde{c} > 0)$. The latter satisfy $H(y, y') = 0$ and are thus the zero-action extremals.

(III) If $-(\beta^2/4) < c < 0$, then $1 < k < \infty$ and if a function $z(t)$ satisfies (3.3) then $w(t) = kz(2\beta^{-1}t)$ satisfies the differential equation

$$w'^2 = (1-w^2)(1-k^{-2}w^2) \tag{3.6}$$

so that $w(t) = \mathrm{sn}(t - \tilde{c}, k^{-1})$. It follows that

$$y(t) = \frac{1}{2} + \frac{1}{2k}\,\mathrm{sn}\left(\frac{\beta}{2}(t - \tilde{c}), \frac{1}{k}\right) \tag{3.7}$$

with period

$$4K = 8\beta^{-1}\int_0^1 ((1 - s^2)(1 - k^{-2}s^2))^{-1/2}\,ds \tag{3.8}$$

and "amplitude" $1/2k$. As k increases from 1 to ∞, this period decreases from ∞ to $4\pi\beta^{-1}$.

(IV) Returning now to the case $c = -\beta^2/4$ [$k = \infty$ in (3.7)] dealt with above, we see that it is natural to think of the solution $y(t) \equiv \frac{1}{2}$ as $\frac{1}{2} + 0\sin(\beta t/2)$ and this will be helpful in our considerations below.

This concludes the summary of extremals but, as is well known, not every extremal minimizes the functional (3.1) and it is easy to see that, unlike the case of mutation studied in [6], there are pairs of points which are joined by more than one extremal lying wholly inside the strip $\Delta = \{(t, y): t \geq 0,\ 0 < y < 1\}$. The following lemma will enable us to identify and reject nonminimizing functions. Note that in what follows we use the term "path" interchangeably with "function" or "graph of function."

LEMMA 3.1. *Suppose that the points $(0, y_0)$ and (T, y_1), where $T > 0$, $0 < y_0 < 1$, $0 < y_1 < 1$, are both on the path (3.7) for given k and $\tilde{c}$. If T is greater than half the period (3.8), then this path does not minimize $S_{0,T}(\phi)$ subject to $\phi(0) = y_0$, $\phi(T) = y_1$.*

PROOF. Let us write $\psi(t)$ instead of $y(t)$ for the function given by (3.7) and let K be defined as in (3.8). Assume first that $\psi(0) \neq \frac{1}{2} \pm 1/2k$. Using the periodicity of $\psi(t)$ and the property $\mathrm{sn}((\beta/2)(2K + x), 1/k) = \mathrm{sn}(-(\beta/2)x, 1/k)$ one can easily check that the extremal

$$\phi(t) = \frac{1}{2} + \frac{1}{2k}\,\mathrm{sn}\left(\frac{\beta}{2}(t + 2K + \tilde{c}), \frac{1}{k}\right)$$

satisfies $\phi(0) = \psi(0)$ and $\phi(2K) = \psi(2K)$. This ϕ is a phase-shifted version of ψ, also passing through the points $(0, \psi(0))$ and $(2K, \psi(2K))$, where clearly $\psi(0) + \psi(2K) = 1$. One of ψ, ϕ is ascending and the other descending at $(0, \psi(0))$. The alternative representation

$$\phi(t) = \frac{1}{2} - \frac{1}{2k}\,\mathrm{sn}\left(\frac{\beta}{2}(2K - \tilde{c} - t), \frac{1}{k}\right)$$

of $\phi(t)$ and the change of variable $t = 2K - u$ in the integral (3.1) easily imply that $S_{0,2K}(\phi) = S_{0,2K}(\psi)$. It follows that if $T > 2K$ and we define

$$\chi(t) = \begin{cases} \phi(t), & \text{if } 0 \leq t \leq 2K, \\ \psi(t), & \text{if } 2K \leq t \leq T, \end{cases}$$

then $S_{0,T}(\chi) = S_{0,T}(\psi)$. If $\psi(t)$, $0 \le t \le T$ were a minimizing function for $S_{0,T}(\cdot)$ subject to the given boundary conditions, then so would χ be. However χ would then be continuously differentiable by Lemma 2.2 and this is not the case since χ is not differentiable at $t = 2K$. This establishes the lemma in the case $\psi(0) \ne \frac{1}{2} \pm 1/2k$. If $\psi(0)$ is either $\frac{1}{2} + 1/2k$ or $\frac{1}{2} - 1/2k$ then $\phi \equiv \psi$ and the above argument has to be modified. Instead of joining the points $(0, \psi(0))$ and $(2K, \psi(2K))$ by an alternative extremal ϕ, we can join the points $(\varepsilon, \psi(\varepsilon))$ and $(\varepsilon + 2K, \psi(\varepsilon + 2K))$ by an alternative extremal, choosing $\varepsilon > 0$ so that $\varepsilon + 2K < T$. We may then argue as above.

4. The preferred paths. We are now in a position to describe the minimizing functions of the functional (3.1) emanating from the point $(0, y_0)$, where $0 < y_0 < 1$. Our description will be much clearer if we do the case $y_0 = \frac{1}{2}$ first. We therefore fix $\phi(0) = \frac{1}{2}$ and present the minimizing paths ϕ corresponding to different values of $\phi'(0)$.

Case (i). If $\phi'(0) = 0$, then the corresponding minimizing path is $\phi(t) \equiv \frac{1}{2}$ but t must be restricted to the domain $0 \le t \le 2\pi\beta^{-1}$.

Case (ii). If $0 < |\phi'(0)| < \beta/4$, then the corresponding paths are

$$\phi(t) = \frac{1}{2} \pm \frac{1}{2k}\,\mathrm{sn}\left(\frac{\beta}{2}t, \frac{1}{k}\right), \tag{4.1}$$

where $k = (\beta/4)|\phi'(0)|^{-1}$ and the sign of the last term is plus or minus according as $\phi'(0) > 0$ or $\phi'(0) < 0$. However, each path of this form must be restricted to a corresponding domain $0 \le t \le 2K$, where K is as in (3.8). Note that $2K$ is the smallest positive value of t at which $\phi(t)$ is equal to $\frac{1}{2}$ and that the two extremals with the same $|\phi'(0)|$ meet at $(2K, \frac{1}{2})$. Both the period and the amplitude of ϕ decrease with increasing k.

Case (iii). If $\phi'(0) = \beta/4$ or $-\beta/4$ then $\phi(t)$ is $(1 + e^{-\beta t})^{-1}$ or $(1 + e^{\beta t})^{-1}$, respectively, and here $0 \le t < \infty$. Note that the last extremal corresponds to "mean behavior" of our Wright–Fisher process.

Case (iv). If $|\phi'(0)| > \beta/4$ then the corresponding paths are

$$\phi(t) = \frac{1}{2} \pm \frac{1}{2}\,\mathrm{sn}\left(\frac{\beta}{2k}t, k\right),$$

where $k = (\beta/4)|\phi'(0)|^{-1}$ and the sign is $+$ or $-$ for $\phi'(0) > 0$ or $\phi'(0) < 0$, respectively. These paths must be drawn from the point $(0, \frac{1}{2})$ to their points of first exit from the strip Δ, that is, the points $(\tilde{K}, 1)$ or $(\tilde{K}, 0)$ of first contact with $y = 1$ or $y = 0$, respectively, where $\tilde{K}$ is a quarter of the period (3.5).

It can be checked that every point (T, y_1) with either $y_1 \ne \frac{1}{2}$ or $0 < T \le 2\pi\beta^{-1}$ lies on exactly one of the paths just described. If $y_1 = \frac{1}{2}$ and $T > 2\pi\beta^{-1}$, then there exist two paths joining $(0, \frac{1}{2})$ with $(T, \frac{1}{2})$ [Case (ii)]. Note the "symmetry" of the minimizing paths, despite the "asymmetry" of genetic selection.

The manner in which Case (ii) extremals were curtailed is justified by Lemma 3.1. To see why Case (iv) extremals must be terminated as indicated, simply note that by Lemma 2.2 a minimizing path joining $(0, \frac{1}{2})$ with (T, y_1) stays clear of the boundary lines $y = 1$ and $y = 0$. Next, observe that the integrals of $H(\phi(t), \phi'(t))$ along the two paths (4.1) from $(0, \frac{1}{2})$ to $(2K, \frac{1}{2})$ are equal, as explained in the proof of Lemma 3.1. There remains to show that if $T > 2\pi\beta^{-1}$, then the constant function $\phi(t) \equiv \frac{1}{2}$ is not a minimizing path joining $(0, \frac{1}{2})$ with $(T, \frac{1}{2})$. In fact, there is a unique k such that $T = 2K$, with K given by (3.8). Let $\phi_1(t)$ be the function (4.1) corresponding to this k, taken with the plus sign. This function satisfies $\phi_1(T) = \frac{1}{2}$. Now choose another k, $k = k_1$ say, so large that the corresponding paths (3.7) drawn beyond their half-period are very close to the line $y = \frac{1}{2}$, with half-period smaller than T. One or the other of these two paths, ψ say, intersects the path ϕ_1 at a point close to $(T, \frac{1}{2})$ with coordinates $(T - \delta, \frac{1}{2} + \delta')$ $(\delta > 0,\ \delta' > 0)$. It is easy to see that $T - \delta$ is greater than half the period of ψ, if $\psi \neq \phi_1$. By Lemma 3.1, ψ cannot minimize the functional S over paths joining $(0, \frac{1}{2})$ with $(T - \delta, \frac{1}{2} + \delta')$. It follows that S is minimized by the extremal joining these points and not crossing the line $y = \frac{1}{2}$ for $0 < t < T - \delta$, that is, by the arc of ϕ_1. Letting $k_1 \to \infty$, we easily deduce $S_{0,T}(\phi_1) \leq S_{0,T}(\phi_{1/2})$, where $\phi_{1/2}(t) = \frac{1}{2},\ 0 \leq t \leq T$. To show that $S_{0,T}(\phi_1) < S_{0,T}(\phi_{1/2})$ suppose by way of contradiction that $S_{0,T}(\phi_1) = S_{0,T}(\phi_{1/2})$ and consider a point $(T_2, \frac{1}{2})$ with $2\pi\beta^{-1} < T_2 < T$. Let ϕ_2 be a Case (ii) path such that $\phi_2(T_2) = \frac{1}{2}$. By what has just been proved $S_{0,T_2}(\phi_2) \leq S_{0,T_2}(\tilde{\phi}_{1/2})$ where $\tilde{\phi}_{1/2}$ is the restriction of $\phi_{1/2}$ to $[0, T_2]$. The function

$$\phi_3(t) = \begin{cases} \phi_2(t), & 0 \leq t \leq T_2, \\ \frac{1}{2}, & T_2 \leq t \leq T \end{cases}$$

would then satisfy $S_{0,T}(\phi_3) \leq S_{0,T}(\phi_{1/2}) \leq S_{0,T}(\phi_1)$, contradicting the non-minimality of ϕ_3 which follows from the fact that ϕ_3 is not differentiable at T_2 (see Lemma 2.2).

The probabilistic theorem that can now be proved is similar to Theorem 2.1, with one difference: if $T > 2\pi\beta^{-1}$ and $y_1 = \frac{1}{2}$, then there are two minimizing paths from $(0, \frac{1}{2})$ to $(T, \frac{1}{2})$. Since the theorem is a special case of Theorem 4.1 below, we do not state it here separately but proceed directly to the analysis of minimizing paths starting from $(0, y_0)$ with $y_0 \neq \frac{1}{2}$. Since all the essential arguments were given for $y_0 = \frac{1}{2}$, we merely outline the picture that emerges.

If $y_0 \neq \frac{1}{2}$ then the constant extremal $\phi(t) \equiv \frac{1}{2}$ plays no role. If the solution $\phi(t)$ of (3.2) is to satisfy $\phi(0) = y_0$, the constant c in (3.2) must be in the range $-\beta^2 y_0(1 - y_0) \leq c < \infty$. Of the extremals represented by (3.7) only functions with "amplitude" $1/2k \geq |y_0 - \frac{1}{2}|$ can possibly pass through $(0, y_0)$. The minimizing paths are then as follows. Assume for convenience that $y_0 > \frac{1}{2}$.

Case (i). If $\phi'(0) = 0$ then the minimizing path $\phi(t)$ is (3.7) with $1/2k = y_0 - \frac{1}{2}$ and $\tilde{c} = -K$, where K is given by (3.8) and with t restricted to $0 \leq t \leq 2K$. We will denote this value of K by K_0.

Case (ii). If $0 < |\phi'(0)| < \beta y_0(1-y_0)$, then the paths are of the form (3.7) with $1/2k > y_0 - \frac{1}{2}$, but must be restricted to $0 \le t \le 2K$ with K given by (3.8). All such ϕ terminate on the line $y(t) \equiv 1 - y_0$ and pairs of minimizing paths with the same $|\phi'(0)|$ (and hence the same period) meet at their terminal points.

Case (iii). If $|\phi'(0)| = \beta y_0(1-y_0)$ then $\phi(t)$ is either $y_0[y_0+(1-y_0)e^{-\beta t}]^{-1}$ or $y_0[y_0+(1-y_0)e^{\beta t}]^{-1}$, the latter representing mean behavior.

Case (iv). If $|\phi'(0)| > \beta y_0(1-y_0)$, then ϕ is of the form (3.4) and must be terminated at its point of exit from Δ.

It should be noted that the role played by the line $y(t) \equiv \frac{1}{2}$ in the case $y_0 = \frac{1}{2}$ is now (i.e., in the case $y_0 > \frac{1}{2}$) played by

$$y(t) = \begin{cases} \phi(t) \text{ as in Case (i)}, & \text{if } 0 \le t \le 2K_0, \\ 1 - y_0, & \text{if } 2K_0 < t < \infty. \end{cases} \tag{4.2}$$

Every point (T, y_1) with either $0 < T \le 2K_0$ or $y_1 \ne 1 - y_0$ lies on exactly one of the paths just described. If $T > 2K_0$ and $y_1 = 1 - y_0$, then there are two Case (ii) paths joining $(0, y_0)$ with (T, y_1) and the integrals of $H(\phi(t), \phi'(t))$ along these two paths from $(0, y_0)$ to (T, y_1) are equal.

The case $y_0 < \frac{1}{2}$ is entirely analogous.

We can now state the main theorem of this paper which is the counterpart, for the case of genetic selection, of Theorem 2.1 already established for the case of mutation.

THEOREM 4.1. (I) *Suppose that* $0 < y_0 < 1$, $T > 0$ *and* $y_0[y_0 + (1 - y_0)e^{\beta T}]^{-1} \le y_1 < 1$ *and that either* $y_1 \ne 1 - y_0$ *or* $T \le 2K_0$, *where* K_0 *is defined as in Case* (i) *above for the extremal* ϕ *with* $\phi(0) = y_0$, $\phi'(0) = 0$. *Let* $\phi_0(T)$, $0 \le t \le T$ *be the unique minimizing path as described above, such that* $\phi_0(0) = y_0$ *and* $\phi_0(T) = y_1$. *Assume that* $Y_0^{(n)} = y_0$ *for all* n *and denote by* P *the probability measure arising from these initial conditions and the transition structure of* $Y_t^{(n)}$, $t \ge 0$ *corresponding to the case of selection* (2.2). *Then, for every* $\delta > 0$,

$$\lim_{n\to\infty} P\Big(\sup_{0\le t\le T} |Y_t^{(n)} - \phi_0(t)| < \delta \mid Y_T^{(n)} \ge y_1\Big) = 1.$$

(II) *With* K_0 *and* P *as in* (I), *suppose that* $y_1 = 1 - y_0$ *and* $T > 2K_0$ *and let* $\phi_1(t), \phi_2(t)$, $0 \le t \le T$ *be the two Case* (ii) *minimizing paths such that* $\phi_1(0) = \phi_2(0) = y_0$ *and* $\phi_1(T) = \phi_2(T) = y_1$. *Then, for every* $\delta > 0$,

$$\lim_{n\to\infty} P\Big(\textit{either} \sup_{0\le t\le T} |Y_t^{(n)} - \phi_1(t)| < \delta \textit{ or } \sup_{0\le t\le T} |Y_t^{(n)} - \phi_2(t)| < \delta | Y_T^{(n)} \ge y_1\Big) = 1.$$

Note that in the special case $y_0 = \frac{1}{2}$, we have $K_0 = 2\pi\beta^{-1}$.

An analogue to part I of the theorem holds if $0 < y_1 \le y_0[y_0+(1-y_0)e^{\beta T}]^{-1}$ and we condition on $Y_T^{(n)} \le y_1$. Part II of the theorem is not relevant in this case because the graph of $y_0[y_0 + (1 - y_0)e^{\beta t}]^{-1}$ lies below the graph of $y(t)$

in (4.2) [this follows from (3.2)], and hence for $T > 2K_0$ we necessarily have $1 - y_0 > y_0[y_0 + (1 - y_0)e^{\beta T}]^{-1}$.

As with Theorem 2.1, we can establish Theorem 4.1 by using the large deviations principle. Since this principle is valid for the modified sequence $\tilde{Y}_{\bullet}^{(n)}$ rather than the sequence $Y_{\bullet}^{(n)}$, we must choose an $\varepsilon > 0$ prior to defining the modification. Such an $\varepsilon > 0$ can be chosen as suggested in the last sentence of Section 2. The application of the large deviations principle then follows the same route as in the proof of Theorem 5.1 in [6], to which the reader is referred.

Acknowledgment. The author is indebted to a referee for some helpful references.

REFERENCES

[1] BOWMAN, F. (1961). *Introduction to Elliptic Functions with Applications*. Dover, New York.

[2] KAPLAN, N. and DARDEN, T. (1985). The behavior of the ratio of a small-noise Markov chain to its deterministic approximation. *Adv. Appl. Probab.* **17** 731–747.

[3] MORROW, G. J. (1992). Large deviation results for a class of Markov chains with applications to an infinite alleles model of population genetics. *Ann. Appl. Probab.* **2** 857–905.

[4] MORROW, G. J. and SAWYER, S. (1989). Large deviation results for a class of Markov chains arising from population genetics. *Ann. Probab.* **17** 1124–1146.

[5] PAPANGELOU, F. (1996). Large deviations of the Wright–Fisher process. *Proceedings of the Athens Conference on Applied Probability and Time Series Analysis 1*: *Applied Probability*. *Lecture Notes in Statist.* **114** 245–252. Springer, New York.

[6] PAPANGELOU, F. (1998). Tracing the path of a Wright–Fisher process with one-way mutation in the case of a large deviation. In *Stochastic Processes and Related Topics: A Volume in Memory of Stamatis Cambanis* (I. Karatzas, B. Rajput and M. S. Taqqu, eds.). Birkhäuser, Boston. To appear.

[7] WENTZELL, A. D. (1990). *Limit Theorems on Large Deviations for Markov Stochastic Processes*. Kluwer, Dordrecht.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MANCHESTER
MANCHESTER M13 9PL
UNITED KINGDOM
E-MAIL: fredos@ma.man.ac.uk