The Annals of Applied Probability
2000, Vol. 10, No. 4, 1259–1273

THE LARGE DEVIATIONS OF A MULTI-ALLELE WRIGHT–FISHER PROCESS MAPPED ON THE SPHERE

BY F. PAPANGELOU

University of Manchester and University of Athens

This is the fourth in a series of papers devoted to the study of the large deviations of a Wright–Fisher process modeling the genetic evolution of a reproducing population. Variational considerations imply that if the process undergoes a large deviation, then it necessarily follows closely a definite path from its original to its current state. The favored paths were determined previously for a one-dimensional process subject to one-way mutation or natural selection, respectively, acting on a faster time scale than random genetic drift. The present paper deals with a general d-dimensional Wright–Fisher process in which any mutation or selection forces act on a time scale no faster than that of genetic drift. If the states of the process are represented as points on a d-sphere, then it can be shown that the position of a subcritically scaled process at a fixed "time" T satisfies a large-deviation principle with rate function proportional to the square of the length of the great circle arc joining this position with the initial one (Hellinger–Bhattacharya distance). If a large deviation does occur, then the process follows with near certainty this arc at constant speed. The main technical problem circumvented is the degeneracy of the covariance matrix of the process at the boundary of the state space.

1. Introduction. It has long been known that the natural geometry of the parameter space of a d-dimensional multinomial distribution is that of the d-sphere. Assuming the size parameter fixed for convenience, if $p = (p_1, \dots, p_{d+1}), q = (q_1, \dots, q_{d+1})$ are two frequency parameter vectors ($p_k \geq 0, q_k \geq 0, \sum_{k=1}^{d+1} p_k = \sum_{k=1}^{d+1} q_k = 1$), the Hellinger–Bhattacharya distance between them is proportional to $\cos^{-1}(\sum_{k=1}^{d+1} \sqrt{p_k q_k})$, which is the angle between the points $(\sqrt{p_1}, \dots, \sqrt{p_{d+1}})$ and $(\sqrt{q_1}, \dots, \sqrt{q_{d+1}})$ on the d-sphere, subtended at the center. See [2] and the historical references given therein; see also [1]. Thus, the orthant of the sphere can be viewed as the natural state space of a d-dimensional Wright–Fisher process modeling the reproduction of a population in which there are $d+1$ alleles.

It is our purpose here to show that the large deviations of a Wright–Fisher process in which mutation and selection act on the same time scale as random genetic drift, in other words, over a number of generations comparable to the size of the population (or possibly more slowly), occur along arcs of great circles of the sphere at uniform speed (Theorem 2) and are governed by a rate function involving the preceding distance in the following sense (Theorem 1): as the size of the population tends to ∞, the position at "time" $T > 0$ of a subcritically

Received July 1998; revised November 1999.
AMS 1991 *subject classifications*. Primary 60F10; secondary 60J20.
Key words and phrases. Wright–Fisher process, random genetic drift, mutation, natural selection, large deviations, rate function, Hellinger–Bhattacharya distance.

scaled Wright–Fisher process starting at p satisfies a large-deviation principle with rate function

$$J_{p,T}(q) = \frac{2}{T}\left[\cos^{-1}\left(\sum_{k=1}^{d+1}\sqrt{p_k q_k}\right)\right]^2. \tag{1.1}$$

Note that the critical scaling is the one leading to the diffusion approximation as explained later.

The problem studied here is of interest for two reasons. First, there are not many examples of non-spatially-homogeneous discrete Markov processes where the rate function for the large deviations of the position can be displayed in closed form. We stress that what we are concerned with here are not the large deviations of either the empirical measure or the sample path of the process but rather those of its position (state) at a "future time." The second cause of interest in the Wright–Fisher process lies in the fact that its covariance matrix is degenerate at the boundary of the state space, where one or more alleles are at or near extinction.

Were it not for this degeneracy, the large-deviation principle for the future position of the process would in principle follow from Wentzell's [16] results at the sample path level via the contraction principle. However, even for processes without degeneracies, the corresponding rate function may not be readily identifiable. One may be able to solve the variational problem and obtain the paths which minimize the action functional between two points as solutions of Euler's equation; it may be harder to calculate the rate function, that is, the minimal value of the action. Even where the latter satisfies the Hamilton–Jacobi equation, this equation may be impossible to solve explicitly. In our present case the identification of the state space with the orthant of a sphere trivializes the variational problem as will be seen.

The degeneracy of the boundary mentioned previously can be handled by means of an inductive argument on the dimension. In the one-dimensional case it was overcome in [11] (cf. also [12] and [13]) by the simple device of modifying the process near the boundary. The higher dimensional case is much more intricate but essentially the inductive step in our proof below exploits the fact that the marginals of a Wright–Fisher process are very nearly lower dimensional Wright–Fisher processes.

It is worth drawing the distinction between the scaled processes of the present paper and those of [12] and [13]. In the latter, mutation and selection were assumed to act on a faster time scale than random genetic drift. This led to much more surprising families of minimizing paths but rate functions are too complicated to calculate explicitly in those cases.

At the end of the paper we comment briefly on two further issues. One concerns the large deviations of the time-scaled diffusion approximation to the Wright–Fisher process. The other concerns the exponential asymptotics of the probability of *reaching* a given subset of the boundary. Boundary sets are not open in the natural topology but a large-deviation lower bound is stated for such sets in the case of pure random genetic drift (Theorem 3).

2. The variational problem. We establish first the notation we will use for our Wright–Fisher process. Suppose that the number of alleles for a gene at a particular locus on a chromosome is $d+1, d \geq 1$. If in a population of $2N$ genes the proportions in which these alleles appear are $i_1/2N, i_2/2N, \dots, i_{d+1}/2N$, $i_1 + i_2 + \cdots + i_{d+1} = 2N$, then the state of the corresponding Wright–Fisher process is the vector $y = (y_1, y_2, \dots, y_d)$, where $y_k = i_k/2N$, $k = 1, 2, \dots, d$. The next generation of $2N$ genes is produced by "sampling with replacement" so that the probability $P(y, \tilde{y})$ of a transition from state $y = (y_1, \dots, y_d)$ to state $\tilde{y} = (j_1/2N, \dots, j_d/2N)$ is given by the following multinomial term (with $j_{d+1} = 2N - \sum_{k=1}^{d} j_k$):

$$P(y, \tilde{y}) = \left[(2N)!\right]\left[\prod_{k=1}^{d+1}(j_k!)\right]^{-1} \pi_1^{j_1}\pi_2^{j_2}\cdots\pi_{d+1}^{j_{d+1}}, \tag{2.1}$$

where in the absence of mutation or selection $\pi_k = y_k$, $k = 1, \dots, d$, and $\pi_{d+1} = 1 - \sum_{k=1}^{d} \pi_k$, while if there are mutation and selection effects the dependence of the π_k's on the y_k's is more involved.

It is a classical fact [5] that if the number of generations over which mutation and selection have an effect is of the same order as the size $2N$ of the population, then, for large N and under appropriate hypotheses, there is a diffusion approximation to the Wright–Fisher process, whose state space is the simplex

$$\Sigma = \left\{(y_1, \dots, y_d) \in \mathbb{R}^d : y_1 \geq 0, \dots, y_d \geq 0 \text{ and } \sum_{k=1}^{d} y_k \leq 1\right\}. \tag{2.2}$$

This diffusion is obtained as a "limit" of the Wright–Fisher process as $N \to \infty$, provided the time axis is scaled so that there are 2N generations in a "unit of time." The Fleming–Viot process introduced in [7] and extensively studied over the last two decades is an extension of such a diffusion to state spaces consisting of probability measures on infinite sets which replace the set $\{1, 2, \dots, d+1\}$.

Here, however, we are primarily interested in the Wright–Fisher process itself rather than its diffusion approximation, although in Section 4 we indicate a companion large-deviation result for the corresponding diffusion process. If the time axis is scaled less severely, so that there are n generations per unit of time, where $n/N \to 0$, then the process converges (over any bounded time interval) to the constant "process" concentrated at the initial state. In particular, its position at "time" T converges in distribution to a point mass situated at the initial position. As mentioned in the Introduction, our purpose is to show that the position at time T satisfies a large-deviation principle.

To define our sequence $Y_t^{(n)}, t \geq 0, n = 1, 2, \dots$, of scaled Wright–Fisher processes, assume that $N_n, n = 1, 2, \dots$, is a sequence of positive integers such that $n/N_n \to 0$ as $n \to \infty$ and write N for N_n for the sake of convenience. For each $n = 1, 2, \dots$ let $Y_0^{(n)}, Y_{1/n}^{(n)}, Y_{2/n}^{(n)}, \dots$ be a d-dimensional Wright–Fisher process with transition probability given by (2.1) with

$$\pi_k = \pi_k(y,n) = y_k + \frac{g_k(y) + o_k(1)}{N}, \qquad k = 1,2,\ldots,d. \tag{2.3}$$

Here it is assumed that $g_1(y), \ldots, g_d(y)$ are continuous functions of $y \in \Sigma$ [see (2.2)] and that $o_k(1) \to 0$ uniformly in y as $n \to \infty$. The process $Y_t^{(n)}$ remains constant on $\nu/n \le t < (\nu+1)/n$ for each $\nu = 0,1,2,\ldots$. Notice that (2.3) differs crucially from the considerations of [12] and [13], where $\pi_1 = y_1 + (g_1(y) + o(1))/n$, although in other respects the analysis of the asymptotics is similar. The cumulant generating function of $Y_t^{(n)}$, $t \ge 0$, in the sense of [16] is $G^n(y,z) = n \log E_y \exp\{z \cdot (Y_{1/n}^{(n)} - y)\}$, where $z \in \mathbb{R}^d$ and $z \cdot y$ denotes the inner product. According to [16], the large deviations of $Y_t^{(n)}$, $n = 1,2,\ldots$, are determined by the asymptotic behavior (as $n \to \infty$) of $n(2N)^{-1}G^n(y,(2N)n^{-1}z)$ and an easy calculation shows that the latter is equal to

$$-n(z\cdot y) + n^2 \log\left\{1 + \sum_{k=1}^{d}\left(\exp\frac{z_k}{n} - 1\right)\pi_k\right\}$$

(see [12] for the case $d = 1$). Using the expansions $\log(1+s) = s - s^2/2 + O(s^3)$ and $e^s - 1 = s + s^2/2 + O(s^3)$, one can show exactly as outlined in [12] that $n(2N)^{-1}G^n(y,(2N)n^{-1}z)$ converges to $G(y,z) = \frac{1}{2}zA(y)z$, where $A(y)$ is the matrix $(a_{kl})_{kl}$, $k,l = 1,\ldots,d$, with $a_{kl} = y_k(\delta_{kl} - y_l)$. (We blur the distinction between z and its transpose.) A number of other conditions required in Wentzell's theory are met (see [12] for the one-dimensional case) but not all, since the quadratic form $zA(y)z$ is not uniformly elliptic, being degenerate on the boundary of Σ. Despite this, for points y that are not on the boundary of Σ, the Legendre transform of $G(y,z)$ is (see[16])

$$H(y,u) = \sup_z\big[z\cdot u - G(y,z)\big] = \frac{1}{2}uA(y)^{-1}u.$$

The matrix $A(y)$ is the covariance matrix of a multinomial random vector and it also provides the Cramér–Rao lower bound (matrix) for the multinomial family of distributions. Its inverse $A(y)^{-1}$, which is the matrix $(b_{kl})_{kl}$, $k,l = 1,\ldots,d$, with $b_{kl} = y_{d+1}^{-1} + \delta_{kl}y_l^{-1}$, where $y_{d+1} = 1 - \sum_{k=1}^{d} y_k$, is the Fisher information matrix for the multinomial distribution. The quadratic form $uA(y)^{-1}u$ defines Rao's Riemannian metric for the family of multinomial distributions, which coincides with the Hellinger–Bhattacharya metric mentioned in the Introduction (see [1] and [2]). To link this with large deviations, we set up here the corresponding variational problem. It can be seen that

$$uA(y)^{-1}u = \sum_{k=1}^{d+1}\frac{u_k^2}{y_k},$$

where $u_{d+1} = -\sum_{k=1}^{d} u_k$. (See, for instance, Lemma 2.1 in [10].) Pretending for a moment that we can ignore the degeneracies, we expect in accordance with Wentzell's theory that, roughly speaking, "away from the boundary" a sample path large-deviation principle is satisfied in the sense that the logarithm of the probability that $Y_t^{(n)}, 0 \le t \le T$ ($T > 0$ fixed), "follows closely" a curve $\phi(t), 0 \le t \le T$, in the interior Σ^o of Σ is of order

$$-2Nn^{-1}\int_0^T H(\phi(t), \phi'(t))dt = -2Nn^{-1}\int_0^T \frac{1}{2}\biggl(\sum_{k=1}^{d+1}\frac{\phi'_k(t)^2}{\phi_k(t)}\biggr)dt,$$

where the integral is taken to be ∞ if ϕ is not absolutely continuous. Thus the functional

$$S_{0,T}(\phi) = \frac{1}{2}\int_0^T \biggl(\sum_{k=1}^{d+1}\frac{\phi'_k(t)^2}{\phi_k(t)}\biggr)dt$$

appears in a role similar to that of an action functional. The probability that a process starting at a point $p \in \Sigma^o$ at time 0 is "at" a point q at time T is roughly of the same exponential order as the probability that the process follows closely the path $\phi(t), 0 \le t \le T$, joining p and q which minimizes the preceding integral. This is essentially the contraction principle. The variational problem to be solved is therefore that of minimizing the integral $\frac{1}{2}\int(\sum_{k=1}^{d+1} y_k'^2/y_k)\,dt$ between the two points. The simple transformation $x_k = \sqrt{y_k}$, $k = 1, 2, \dots, d+1$, converts this to the problem of minimizing the action $2\int(\sum_{k=1}^{d+1} x_k'^2)\,dt$ subject to $\sum_{k=1}^{d+1} x_k^2 = 1$. The extremals (see [8]) of this problem satisfy $\sum_{k=1}^{d+1} x_k'^2 = c^2$ (a constant) and are thus geodesics (great circles) of the sphere $\sum_{k=1}^{d+1} x_k^2 = 1$ with t proportional to arc length.

Thus if $p = (p_1, \dots, p_d)$ and $q = (q_1, \dots, q_d)$ are two points of Σ^o and we set $J_{p,T}(q) =: \inf\{S_{0,T}(\phi) : \phi(t), 0 \le t \le T$, is a curve in Σ^o with $\phi(0) = p$ and $\phi(T) = q\}$, then $J_{p,T}(q)$ is equal to

$$2c^2T = 2\biggl(\frac{\theta}{T}\biggr)^2 T = \frac{2}{T}\Bigl[\cos^{-1}(\sqrt{p_1q_1} + \sqrt{p_2q_2} + \cdots + \sqrt{p_{d+1}q_{d+1}})\Bigr]^2,$$

where $p_{d+1} = 1 - \sum_{k=1}^{d} p_k$, $q_{d+1} = 1 - \sum_{k=1}^{d} q_k$ and θ is the "angle" between the images of p and q on the d-sphere.

It is worth mentioning here that the square of $\cos^{-1}(\sum_{k=1}^{d+1}\sqrt{p_kq_k})$ was proposed by Bhattacharya as a "measure of divergence" between two multinomial distributions. In the context of a one-dimensional Wright–Fisher process the angular distance goes back to Fisher's "angular transformation." See [6] and [9]. The metric $\cos^{-1}(\sum_{k=1}^{d+1}\sqrt{p_kq_k})$, which can be defined for boundary points as well, was extended to the infinite-dimensional context of a Fleming–Viot process by Schied [15], who showed that it is the intrinsic metric arising from the "carré du champs" operator of the process.

In the next section we show that, notwithstanding the degeneracy of $A(y)$, $J_{p,T}(q)$ in (1.1) is the rate function of a large-deviation principle for $Y_T^{(n)}$.

3. The large deviations of $Y_T^{(n)}$. Throughout the present paper the topology of Σ considered is the relative topology of Σ in $\mathbb{R}^d$. We have already denoted the interior of Σ by Σ^o. If $p \in \Sigma^o$, denote by P_p the probability measure arising from the initial conditions $Y_0^{(n)} = p$ and the transition structure of $Y_t^{(n)}$, $t \geq 0$, indicated in the preceding section.

THEOREM 1. *Let K be a closed subset of Σ^o, F a closed subset of Σ and G a subset of Σ open in the relative topology of Σ. Then, for any $T > 0$,*

$$\limsup_{n\to\infty} \frac{n}{2N} \log \sup_{p\in K} P_p(Y_T^{(n)} \in F) \leq -\inf_{p\in K} \inf_{q\in F} J_{p,T}(q), \tag{3.1}$$

$$\liminf_{n\to\infty} \frac{n}{2N} \log \inf_{p\in K} P_p(Y_T^{(n)} \in G) \geq -\sup_{p\in K} \inf_{q\in G} J_{p,T}(q). \tag{3.2}$$

The proof will take up most of this section.

To bypass the degeneracy, we follow the practice of [12] and [13] and modify the process near the boundary of Σ in $\mathbb{R}^d$ by choosing a suitable neighborhood Σ' of the boundary in Σ and replacing the distribution of the step of our process out of a point $y' \in \Sigma'$ by a Gaussian step with mean $N^{-1}(g_1(y), \dots, g_d(y))$ and covariance matrix $N^{-1}A(y)$ [or $N^{-1}A(\pi)$], where y is the point of $\Sigma\backslash\Sigma'$ nearest y'. The quadratic form of the new covariance matrix is bounded away from 0 so that Wentzell's results apply (Theorem 3.2.3$'$ in [16]). See [12] for the one-dimensional case. We will denote by $\widetilde{Y}_t^{(n)}$, $t \geq 0$, the modified process, with state space $\mathbb{R}^d$. The proof of the lower bound (3.2) is now straightforward. Suppose that $G \subset \Sigma$ is open in the relative topology of Σ, $K \subset \Sigma^o$ is compact and $q \in G \cap \Sigma^o$. We can choose a closed neighborhood Σ' of the boundary of Σ such that $\Sigma' \cap K = \emptyset$, $q \notin \Sigma'$ and the geodesics (with respect to the metric introduced in the preceding section) joining points of K with q lie entirely in $\Sigma\backslash\Sigma'$. Suppose $p \in K$ and let ϕ_0 be the geodesic joining p with q. If U denotes the set of curves $\phi(t)$, $0 \leq t \leq T$, in Σ such that ϕ does not exit from $\Sigma\backslash\Sigma'$ and $\phi(T) \in G$, then, by the large-deviation principle for the sample paths of the modified process $\widetilde{Y}^{(n)} = \{\widetilde{Y}_t^{(n)}, 0 \leq t \leq T\}$,

$$\begin{aligned}
\liminf_{n\to\infty} \frac{n}{2N} \log P_p(Y_T^{(n)} \in G) &\geq \liminf_{n\to\infty} \frac{n}{2N} \log P_p(\widetilde{Y}^{(n)} \in U) \\
&\geq -\inf\{S_{0,T}(\phi) : \phi \in U\} \\
&= -S_{0,T}(\phi_0) \\
&= -J_{p,T}(q).
\end{aligned}$$

Continuity arguments establish the same for q on the boundary of Σ and it is also a simple matter to extend this to the whole set K of initial states using the uniformity known for $\widetilde{Y}^{(n)}$.

The upper bound is more delicate because we need to handle sets of paths that can get arbitrarily close to the boundary. In this respect Azencott's [3] general methods at sample path level are not helpful in our case because of the nature of the topology on paths on which they are based.

LEMMA 3.1. *Given any $M > 0$ and $\eta > 0$, there is a $\delta > 0$ such that, for any $p \in \Sigma^o$,*

$$\limsup_{n\to\infty} \frac{n}{2N} \log P_p\Big\{ \text{there exist } t', t'' \in [0, T] \text{ such that } |t' - t''| < \delta \text{ and } \|Y_{t'}^{(n)} - Y_{t''}^{(n)}\| \geq \eta \Big\} \leq -M,$$

where $\|\cdot\|$ denotes the Euclidean norm.

PROOF. The proof is an adaptation of arguments given in [16]. It is sufficient to establish the lemma for each coordinate of $Y_t^{(n)}$ and we do so for the first one which we denote by $Y_{1,t}^{(n)}$. If $y = (y_1, y_2, \dots, y_d)$ and $\tilde{z} = (z, 0, 0, \dots, 0)$, where $z \in \mathbb{R}$, then the cumulant generating function is

$$G^n(y, \tilde{z}) = -nzy_1 + 2nN \log\bigg\{1 + \pi_1\bigg(\exp\frac{z}{2N} - 1\bigg)\bigg\},$$

where π_1 is defined in (2.3). It is easy to show that if C is an upper bound of $|g_1(y) + o(1)|$ in (2.3), then $G^n(y, \tilde{z}) \leq \overline{G}^n(z)$, where

$$\overline{G}^n(z) = \frac{n}{N}\bigg[C|z| + \frac{1}{4}\bigg(1 + \frac{C}{N}\bigg)z^2\bigg].$$

(If $z > 0$, write $G^n(y, \tilde{z})$ in the form $nz(1 - y_1) + 2nN \log\{1 + (1 - \pi_1)(\exp(-z/2N) - 1)\}$ and use the inequality $\log(1 + x) \leq x$. The case $z < 0$ is more direct.) The fact that the process

$$\exp\bigg\{z(Y_{1,m/n}^{(n)} - Y_{1,l/n}^{(n)}) - \frac{1}{n}\sum_{i=l}^{m-1} G^n(Y_{i/n}^{(n)}, \tilde{z})\bigg\}, \qquad m = l+1, l+2, \dots,$$

is a martingale, combined with the inequality just stated, implies that if $t' < t''$, then

$$E\Big\{\exp\Big(z(Y_{1,t''}^{(n)} - Y_{1,t'}^{(n)})\Big)|Y_{t'}^{(n)} = y\Big\} \leq \exp\Big\{(t'' - t')\overline{G}^n(z)\Big\}.$$

A standard large-deviation argument shows that if $0 < t'' - t' < \delta$ and $u > 0$, then

$$P_p\Big\{Y_{1,t''}^{(n)} - Y_{1,t'}^{(n)} \geq \delta u | Y_{t'}^{(n)} = y\Big\} \leq \exp\Big\{\delta\Big(\overline{G}^n(z) - zu\Big)\Big\}$$

for any y and z. If, given $M > 0$ and $\eta_0 > 0$, we set $\delta = \eta_0^2(4M)^{-1}$, $u = \eta_0\delta^{-1}$ and $z = Nn^{-1}u$, we see that $0 < t''-t' < \delta$ implies that, for sufficiently large n,

$$P_p\Big\{Y_{1,t''}^{(n)} - Y_{1,t'}^{(n)} \geq \eta_0 | Y_{t'}^{(n)} = y\Big\} \leq \exp\Big\{-2NMn^{-1}\Big\}.$$

The case $Y^{(n)}_{1,t''} - Y^{(n)}_{1,t'} \le -\eta_0$ can be handled similarly. If now $0 = t_0 < t_1 < \cdots < t_m = T$ is a partition of $[0, T]$ such that $\max_{1\le\nu\le m}(t_\nu - t_{\nu-1}) < \delta$, then one may use arguments similar to those on pages 42–43 of [16] to show that

$$P_p\Big\{|Y^{(n)}_{1,t} - Y^{(n)}_{1,t_{\nu-1}}| > 2\eta_0 \text{ for some } t \in [t_{\nu-1}, t_\nu]\Big\}$$
$$\le 2 \sup_{\substack{t_{\nu-1}\le t\le t_\nu \\ y\in\Sigma^o}} P\Big\{|Y^{(n)}_{1,t_\nu} - Y^{(n)}_{1,t}| \ge \eta_0 | Y^{(n)}_t = y\Big\}$$

from which the lemma follows easily.

The large-deviation upper bound will be proved by induction on the dimension d of the Wright–Fisher process. If $d = 1$ the result follows immediately from the large-deviation upper bound for the suitably modified process $\widetilde{Y}^{(n)}$, since it is sufficient to consider closed sets F of the form $F = [0, q'] \cup [q, 1]$, where $q' < p < q$. Briefly, if the scaled process reaches q say at time T, having previously approached 1, then it must have gone through q at an earlier time. The probability of this, however, is of a lower order than the probability that it has reached q for the "first time" at time T. The full argument was given in [11] for the case of pure random genetic drift [$\pi = y$ in (2.3)] and the rate function was produced there in the form of an integral which can be calculated trivially. It extends to the present case because of the assumed rate of action of mutation and selection. The situation differs from that of [12] and [13], where the action of mutation or selection is faster.

The inductive argument will exploit the fact that the marginal processes of a Wright–Fisher process are very nearly Wright–Fisher processes: whenever the process gets close to the boundary of Σ we drop down to a lower dimension and observe only the marginal process. The idea is to construct a "fence" just inside the boundary of Σ in such a way that Hellinger–Bhattacharya distances along the fence are equal to distances traveled by the marginal in its own state space. This will become clear in the following paragraphs.

To avoid complicating the issue and cluttering the notation, we will outline the proof for $d = 2$, omitting details wherever the arguments are obvious. Except for the last paragraph, throughout this proof Σ will denote the set $\{p = (p_1, p_2)\colon p_1 \ge 0, p_2 \ge 0, p_1 + p_2 \le 1\}$ and $\rho(p, q)$ the Hellinger–Bhattacharya distance between two points $p = (p_1, p_2)$ and $q = (q_1, q_2)$ of Σ:

$$\rho(p, q) = \cos^{-1}\left(\sqrt{p_1 q_1} + \sqrt{p_2 q_2} + \sqrt{(1 - p_1 - q_1)(1 - p_2 - q_2)}\right).$$

We will assume for convenience that $K = \{p\}$, where $p \in \Sigma^o$. Let F be a closed subset of Σ such that $p \notin F$. Let β be an arbitrary positive number less than $\inf\{\rho(p, r)\colon r \in F\}$ and cover each point $q \in F$ by a compact neighborhood F_q of ρ-diameter less than $\beta/2$ and also less than half the distance from q to the boundary of Σ if, in addition, $q \in \Sigma^o$. Since F is covered by a finite collection of such neighborhoods we will concentrate on one of them, F_q say, which we will denote by F for convenience. Assume first that $q \in \Sigma^o$. Define $\zeta = \beta/I$, where I is an integer to be determined later.

Note first that straight lines passing through the point $(0, 1)$ (i.e., with equations of the form $ay_1 + y_2 = 1$ or, equivalently, $\sqrt{a-1} \cdot \sqrt{y_1} - \sqrt{1 - y_1 - y_2} = 0$, $a > 1$) are geodesics of the ρ-distance, that is, great circles on the 2-sphere of points $(\sqrt{y_1}, \sqrt{y_2}, \sqrt{1 - y_1 - y_2})$. The same applies to lines through $(1, 0)$ or $(0, 0)$.

Draw the four straight lines $by_1 + y_2 = 1$, $(1 + 1/b)y_1 + y_2 = 1$, $y_1 + by_2 = 1$, $y_1 + (1+1/b)y_2 = 1$ with b so large that the quadrilateral bounded by them contains in its interior the point p, the set $F = F_q$ and the "geodesic" joining p and q. The next step is to "thicken" the boundary of this quadrilateral by drawing four lines $b^* y_1 + y_2 = 1$, $(1 + 1/b^*)y_1 + y_2 = 1$, $y_1 + b^* y_2 = 1$, $y_1 + (1 + 1/b^*)y_2 = 1$ with b^* slightly larger than b and chosen in a manner to be described later. Denote by Q_1^o the open quadrilateral enclosed by the first set of lines and by $\overline{Q}_2$ the closed (and larger) quadrilateral enclosed by the second set of lines. The set $\overline{Q}_2 \backslash Q_1^o$ will play the role of the "fence" referred to previously. Next denote by R_1 the set of points $y = (y_1, y_2)$ of $\overline{Q}_2 \backslash Q_1^o$ which lie between the lines $b^* y_1 + y_2 = 1$ and $by_1 + y_2 = 1$, that is, $R_1 = \{(y_1, y_2) \in \overline{Q}_2 \backslash Q_1^o : by_1 + y_2 < 1 < b^* y_1 + y_2\}$ and let R_2, R_3, R_4, be the other three thick "sides" making up the fence, defined similarly. Note that these sides are not disjoint.

In our construction we choose b^* so close to b that the set $\overline{Q}_2 \backslash Q_1^o$ can be expressed as a finite union $\cup_{k=1}^{L} D_k$ of small compact sets D_k, $k = 1, 2, \ldots, L$, having the following properties:

(Π_1) If D_k intersects any one of the sets R_1, R_2, R_3, R_4, then it is contained in it.

(Π_2) If a point $y = (y_1, y_2)$ of R_1 is in D_k, then every point $y' = (y_1', y_2)$ of R_1 with the same ordinate y_2 is within ρ-distance less than ζ from every point of D_k. Likewise with R_2, R_3, R_4.

Next define

$$\eta = \frac{1}{2} \inf\Big\{ ||y - y'|| : y \in Q_1^o,\ y' \notin \overline{Q}_2 \Big\},$$

fix an $M > (2/T)\rho(p, q)^2$ and let $\delta > 0$ be as in Lemma 3.1.

We now express the set $W^{(n)}$ of paths of $Y^{(n)}$ with $Y_0^{(n)} = p$, $Y_T^{(n)} \in F$ as a finite union of sets $W_0^{(n)}, W_1^{(n)}, \ldots$ as follows. We take $W_0^{(n)}$ to be the set of paths such that $||Y_{t'}^{(n)} - Y_{t''}^{(n)}|| \geq \eta$ for some pair t', t'' with $|t' - t''| < \delta$. This set can be handled by means of Lemma 3.1, which implies that

$$P_p(W_0^{(n)}) \leq \exp\Big\{ -\frac{2N}{n} \cdot \frac{2}{T} \rho(p, q)^2 \Big\}$$

for sufficiently large n. The next set, $W_1^{(n)}$, is the set of all paths which remain within Q_1^0. This set can be handled by means of the large-deviation property of the suitably modified process $\widetilde{Y}^{(n)}$ as indicated earlier. Accordingly,

$$P_p(W_1^{(n)}) \le \exp\left\{-\frac{2N}{n}\cdot\frac{2}{T}[\rho(p,q)-\beta]^2\right\}$$

eventually.

To handle the remaining paths, we introduce a partition $0 = t_0 < t_1 < \cdots < t_m = T$ of $[0,T]$ such that $\max_{1\le\nu\le m}(t_\nu - t_{\nu-1}) < \delta$. Every path in $W^{(n)}$ which is not in $W_0^{(n)}$ or $W_1^{(n)}$ necessarily visits $\overline{Q}_2\backslash Q_1^o$ at one or more of the times $t_1, t_2, \dots, t_{m-1}$. Considering only these times, we can decompose the set of remaining paths according to the order in which any of R_1, R_2, R_3, R_4 are visited and also according to the particular D_k's that are visited. To take an example, suppose $D_k \subset R_1, D_l \subset R_1$ and $0 < i < j < m$, and let $W_2^{(n)}$ be the set of paths of $Y_t^{(n)}, 0 \le t \le T$, such that $Y_t^{(n)} \in Q_1^0$ for $0 \le t \le t_{i-1}$, $Y_{t_i}^{(n)} \in D_k$, $Y_{t_j}^{(n)} \in D_l$ and $Y_t^{(n)} \in Q_1^o$ for $t_{j+1} \le t \le T$. Then

$$\begin{aligned} P_p(W_2^{(n)}) \le \int_{y'\in D_l}\int_{y\in D_k} & P_p\{Y_t^{(n)} \in \overline{Q}_2 \text{ for } 0\le t\le t_i, Y_{t_i}^{(n)} \in dy\} \\ &\times P\{Y_{t_j}^{(n)} \in dy' | Y_{t_i}^{(n)} = y\} \\ &\times P\{Y_s^{(n)} \in \overline{Q}_2 \text{ for } t_j \le s \le T, Y_T^{(n)} \in F | Y_{t_j}^{(n)} = y'\}. \end{aligned} \tag{3.3}$$

Given any $\theta \in (0,1)$, the last probability in the double integral is, for all $y' \in D_l$, less than

$$\begin{aligned} &\exp\Big[-\frac{2N}{n}(1-\theta)\inf\{S_{t_j,T}(\phi) : \phi(t_j) \in D_l, \phi(s) \in \overline{Q}_2 \\ &\qquad\qquad \text{for all } s \in [t_j, T], \phi(T) \in F\}\Big] \\ &\quad \le \exp\left[-\frac{2N}{n}(1-\theta)\inf\left\{2(T-t_j)^{-1}\rho(y',y'')^2 : y' \in D_l, y'' \in F\right\}\right] \end{aligned} \tag{3.4}$$

eventually. Next, if $Z_t^{(n)}$ denotes the second coordinate of $Y_t^{(n)}$ and $\widetilde{D}_l = \{y_2 : (y_1, y_2) \in D_l \text{ for some } y_1\}$, then $P\{Y_{t_j}^{(n)} \in D_l | Y_{t_i}^{(n)} = y\} \le P\{Z_{t_j}^{(n)} \in \widetilde{D}_l | Y_{t_i}^{(n)} = y\}$. The process $Z_t^{(n)}$, $t \ge 0$, is not Markovian under its marginal distribution, since π_2 in (2.3) may depend on both y_1 and y_2. However, $Z_t^{(n)}$ can be sandwiched between two scaled one-dimensional Wright–Fisher processes, $Z_t'(n), Z_t''(n)$ say, one with π_2' of the form $y_2 - C/N$ suitably truncated at 0 and one with π_2'' of the form $y_2 + C/N$ suitably truncated at 1, where C is an upper bound of $|g_2(y) + o_2(1)|$ in (2.3). [It is sufficient to note that if X', X'' are random variables with binomial distributions $B(2N, \pi')$ and $B(2N, \pi'')$, respectively, where $\pi' \le \pi''$, then X' and X'' can be "coupled" so that $X' \le X''$ almost surely. This may be achieved for instance by "thinning" the successes in a sequence of Bernoulli trials.] Since each of the sequences $Z_{t_j}'(n)$ and $Z_{t_j}''(n)$

satisfies a large-deviation principle, it is easy to see that we have

$$\begin{aligned}&\sup_{y\in D_k} P\Big\{Y^{(n)}_{t_j}\in D_l \big| Y^{(n)}_{t_i}=y\Big\}\\ &\quad\le \exp\Big[-\frac{2N}{n}(1-\theta)\inf\Big\{2(t_j-t_i)^{-1}\tilde\rho(y_2,y_2')^2 : y_2\in\widetilde D_k,\, y_2'\in\widetilde D_l\Big\}\Big]\end{aligned}\tag{3.5}$$

eventually, where $\tilde\rho(y_2,y_2')=\cos^{-1}(\sqrt{y_2y_2'}+\sqrt{(1-y_2)(1-y_2')})$. Now we know that the point $y=(y_1,y_2)$ is in $D_k\subset R_1$. Consider the straight line through (y_1,y_2) and $(0,1)$. For $y_2'\in\widetilde D_l$ there is a y_1'' such that the point (y_1'',y_2') is on this line and the crucial argument is that, as can be checked,

$$\sqrt{(1-y_2)(1-y_2')}=\sqrt{y_1y_1''}+\sqrt{(1-y_1-y_2)(1-y_1''-y_2')}$$

so that $\tilde\rho(y_2,y_2')=\rho(y,y'')$, where $y''=(y_1'',y_2')$. By condition (Π_2) the right-hand side of (3.5) is less than

$$\exp\Big[-\frac{2N}{n}(1-\theta)\inf\Big\{2(t_j-t_i)^{-1}\big(\rho(y,y')-\zeta\big)^2 : y\in D_k,\, y'\in D_l\Big\}\Big].\tag{3.6}$$

Finally,

$$\begin{aligned}&P_p\Big\{Y^{(n)}_t\in\overline{Q}_2 \text{ for } 0\le t\le t_i,\, Y^{(n)}_{t_i}\in D_k\Big\}\\ &\quad\le \exp\Big[-\frac{2N}{n}(1-\theta)\inf\Big\{2t_i^{-1}\rho(p,y)^2 : y\in D_k\Big\}\Big]\end{aligned}\tag{3.7}$$

eventually.

Since the ρ-diameter of every D_k is less than $\zeta=\beta/I$, we see that if $I>5$ then $\rho(p,q)-\beta$ is less than $\inf\rho(p,y)+\inf(\rho(y,y')-\zeta)+\inf\rho(y',y'')$, where the three infima are taken over the ranges indicated in the right-hand sides of (3.7), (3.6) and (3.4), respectively. Using the fact that

$$\frac{(x_1+x_2+x_3)^2}{T_1+T_2+T_3}\le\frac{x_1^2}{T_1}+\frac{x_2^2}{T_2}+\frac{x_3^2}{T_3}$$

if $x_i>0$, $T_i>0$ (Cauchy–Schwarz), we see from (3.3) that eventually

$$P_p\big(W_2^{(n)}\big)\le\exp\Big\{-\frac{2N}{n}(1-\theta)\frac{2}{T}\big[\rho(p,q)-\beta\big]^2\Big\}.\tag{3.8}$$

This procedure for handling the set $W_2^{(n)}$ of paths actually shows that in this expression $\rho(p,q)$ can be replaced by the infimum of ρ-lengths of curves $\phi(t)$, $0\le t\le T$, in $\overline{Q}_2$ which start at p, exit from Q_1^o and end at points of F.

One may next consider paths for which the sequence $Y^{(n)}_{t_1}, \ldots, Y^{(n)}_{t_m}$, after leaving R_1 for the last time visits R_2, say. We then need to consider the last time it leaves R_2, and so on. If between the last time it leaves R_1 and the first subsequent time it visits R_2 the path travels outside $\overline{Q}_2$ and negotiates around $R_1 \cap R_2$ say, then we need to consider first one of its coordinates (marginals) and then the other. The length of the great circle arc joining the endpoints of this stretch of the path is no greater than a sum of arc lengths along the sides of the fence. The arguments are similar to those given previously, the only adjustment required being in the value of I in the definition of $\zeta = \beta/I$.

The large-deviation upper bound for $F = F_q$ follows from the fact that the decomposition of paths outlined is finite.

Cases in which the point q is on the boundary of Σ can be treated by analogous arguments. If, for instance, such a q is not a vertex of Σ, then the fence is constructed in such a way that one of the sets D_k [see (Π_1) and (Π_2)] is contained in the neighborhood F_q. Each path starting at p and terminating in F_q is then split as before into various sections, corresponding to the movements of the path in and out of the fence. If now such a path terminates outside $\overline{Q}_2$, then the final section to be considered will have been shadowed by a "marginal" process along the fence.

Finally, we comment briefly on the inductive step for higher dimensions d. We have shown how to move up from $d = 1$ to $d = 2$. To illustrate the general step from d to $d+1$, we consider the three-dimensional Wright–Fisher process whose state space is the tetrahedron $\Sigma = \{(y_1, y_2, y_3) : y_1 \geq 0, y_2 \geq 0, y_3 \geq 0, y_1 + y_2 + y_3 \leq 1\}$. In this case the fence is constructed with the aid of planes, each of which contains one of the edges of the tetrahedron. If $p = (p_1, p_2, p_3)$ and $q = (q_1, q_2, q_3)$ are two points on such a plane, then it can be checked that the Hellinger–Bhattacharya distance between p and q is equal to the Hellinger–Bhattacharya distance between their "marginals," (p_1, p_2) and (q_1, q_2) say. As regards the "sandwiching" of a two-dimensional marginal, this is achieved by means of four two-dimensional Wright–Fisher processes. It should now be clear how the inductive arguments proceed.

This concludes the proof of Theorem 1.

The arguments establishing Theorem 5.1 in [12], combined with the remark immediately after (3.8), easily imply the following result.

THEOREM 2. *Suppose $p \in \Sigma^o$, $q \in \Sigma^o$ and let $\phi_0(t)$, $0 \leq t \leq T$, be the curve in Σ^o whose image on the d-sphere is the arc of the great circle joining the images of p and q on the sphere, with parameter t proportional to arc length. Let $\delta > 0$. If G is a sufficiently small neighborhood of q, then*

$$\lim_{n\to\infty} P_p\Big\{\sup_{0\leq t\leq T} ||Y^{(n)}_t - \phi_0(t)|| < \delta | Y^{(n)}_T \in G\Big\} = 1.$$

We omit the purely technical extension to $q \in \Sigma$. The theorem roughly asserts that if $Y^{(n)}_t$, $0 \leq t \leq T$, undergoes a large deviation from $p \in \Sigma^o$ to $q \in \Sigma^o$, then, with near certainty, its image on the d-sphere follows closely the curve which traverses the arc of the great circle joining the images of p and q, at constant speed.

4. Two remarks.

REMARK 1. Results entirely analogous to Theorems 1 and 2 can be proved for the continuous-time version of the Wright–Fisher process, that is, the so-called "diffusion approximation." Consider a degenerate diffusion Y_t, $t \geq 0$, on Σ with differential operator of the form

$$\frac{1}{2}\sum_{k=1}^{d}\sum_{l=1}^{d} y_k(\delta_{kl} - y_l)\frac{\partial^2}{\partial y_k \partial y_l} + \sum_{k=1}^{d} b_k(y_1, \dots, y_d)\frac{\partial}{\partial y_k}.$$

For $\varepsilon > 0$ define $Y_t^\varepsilon = Y_{\varepsilon t}$, $t \geq 0$. Under mild conditions on the b_k's and for fixed $T > 0$, Y_T^ε satisfies a large-deviation principle as $\epsilon \to 0$, with rate function given by (1.1), that is,

$$\limsup_{\varepsilon \to 0} \varepsilon \log P_p(Y_T^\varepsilon \in F) \leq -\inf_{q \in F} J_{p,T}(q),$$
$$\liminf_{\varepsilon \to 0} \varepsilon \log P_p(Y_T^\varepsilon \in G) \geq -\inf_{q \in G} J_{p,T}(q),$$

where $p \in \Sigma^o$ and F, G are as in Theorem 1. Here again this does not follow from Theorem 6.4 in [3]. If $Y_\bullet^\varepsilon$ undergoes a large deviation, then its image on the sphere traces an arc of a great circle, in the sense of Theorem 2.

Just as in the discrete case of Section 3, the scaling $Y_t^\varepsilon = Y_{\varepsilon t}$ has the effect of scaling down the three factors of random genetic drift, mutation and selection at the same rate. This differs somewhat from the more usual scaling which is applied to the diffusion coefficient only, as is, for instance, the case with Theorem 3.3 in [4]. This latter theorem (which appeared after the first version of the present paper was submitted) establishes a large-deviation principle at path level for a class of degenerate diffusions such as the above but does not cover our case: it deals with a case of unscaled neutral mutation in which the rate function diverges to ∞ as the path approaches the boundary of Σ. This feature eliminates the need to consider the behavior of a path beyond its initial approach to the boundary.

REMARK 2. The degeneracy of a Wright–Fisher process at the boundary raises an interesting problem, namely that of determining the rough asymptotics of the probabilities of actually reaching given subsets of the boundary of Σ. Such subsets are not open in the relative topology of Σ and hence the large-deviation lower bound of Theorem 1 can only tell us something about the probability of getting "close" to them. However, the event of actually reaching the boundary is of great importance. In the absence of mutation, for instance, the phenomenon of fixation, whereby all but one of the alleles eventually become extinct, is of primary interest. Theorem 3 provides a lower bound for subsets of the boundary under the additional assumptions that $n^2 = o(N)$ and that the Wright–Fisher process is subject only to random genetic drift; that is, there is no mutation or selection. The proof of the theorem is given in [14]. The large-deviation upper bound for boundary sets follows trivially from Theorem 1.

THEOREM 3. *Let* $\Sigma_m = \{q = (q_1, \ldots, q_d) \in \Sigma : q_1 = q_2 = \cdots = q_m = 0\}$, *where* $m \leq d$, *and suppose that* G *is a subset of* Σ_m *which is open in the relative topology of* Σ_m. *If* $Y_t^{(n)}, 0 \leq t \leq T$, *is defined as in Section 2 but with* $\pi_k = y_k$ *in* (2.3) *and if* $n^2 = o(N)$ *as* $n \to \infty$, *then, for* $p \in \Sigma^o$,

$$\liminf_{n\to\infty} \frac{n}{2N} \log P_p(Y_T^{(n)} \in G) \geq -\inf_{q\in G} J_{p,T}(q).$$

The theorem naturally holds for all other components of the boundary of Σ as well.

Acknowledgment. The author is indebted to a referee for several valuable comments which have led to a better presentation.

REFERENCES

[1] AMARI, S. (1985). *Differential-Geometrical Methods in Statistics. Lecture Notes in Statist*. **28**. Springer, Berlin.

[2] ATKINSON, C. and MITCHELL, A. F. (1981). Rao's distance measure. *Sankhyā Ser. A* **43** 345–365.

[3] AZENCOTT, R. (1980). *Grandes Déviations et Applications. Lecture Notes in Math*. **774** 1–176. Springer, Berlin.

[4] DAWSON, D. A. and FENG, S. (1998). Large deviations for the Fleming–Viot process with neutral mutation and selection. *Stochastic Processes Appl*. **77** 207–232.

[5] EWENS, W. J. (1979). *Mathematical Population Genetics.* Springer, New York.

[6] FISHER, R. A. (1922). On the dominance ratio. *Proc. Roy. Soc. Edinburgh* **42** 321–341.

[7] FLEMING, W. H. and VIOT, M. (1979). Some measure-valued Markov processes in population genetics theory. *Indiana Univ. Math. J*. **28** 817–843.

[8] GELFAND, I. M. and FOMIN, S. V. (1963). *Calculus of Variations*. Prentice-Hall, Englewood Cliffs, NJ.

[9] KIMURA, M. (1964). *Diffusion Models in Population Genetics.* Methuen, London.

[10] 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.

[11] PAPANGELOU, F. (1996). Large deviations of the Wright–Fisher process. *Lecture Notes in Statist*. **114** 245–252. Springer, New York.

[12] 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.) 315–330. Birkhauser, Boston.

[13] PAPANGELOU, F. (1998). Elliptic and other functions in the large deviations behavior of the Wright–Fisher process. *Ann. Appl. Probab.* **8** 182–192.

[14] PAPANGELOU, F. (2000). A note on the probability of rapid extinction of alleles in a Wright–Fisher process. In *Probability and Statistical Models with Applications–A Volume in Honor of T. Cacoullos* (Ch. A. Charalambides, M. V. Koutras and N. Balakrishnan, eds.) 147–154. Chapman and Hall, New York.

[15] SCHIED, A. (1997). Geometric aspects of Fleming–Viot and Dawson–Watanabe processes. *Ann. Probab.* **25** 1160–1179.

[16] 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

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ATHENS
PANEPISTEMIOPOLIS
157 84 ATHENS
GREECE