

HUXLEY AND FISHER EQUATIONS FOR GENE PROPAGATION: AN EXACT SOLUTION

P. BROADBRIDGE¹, B. H. BRADSHAW¹, G. R. FULFORD² and G. K. ALDIS³

(Received 28 April, 2000; revised 13 September, 2000)

Abstract

The derivation of gene-transport equations is re-examined. Fisher's assumptions for a sexually reproducing species lead to a Huxley reaction-diffusion equation, with cubic logistic source term for the gene frequency of a mutant advantageous recessive gene. Fisher's equation more accurately represents the spread of an advantaged mutant strain within an asexual species. When the total population density is not uniform, these reaction-diffusion equations take on an additional non-uniform convection term. Cubic source terms of the Huxley or Fitzhugh-Nagumo type allow special nonclassical symmetries. A new exact solution, not of the travelling wave type, and with zero gradient boundary condition, is constructed.

1. Introduction

At the end of the 20th century, the subject of population genetics has been given additional impetus by the imperatives of new technological developments in bioengineering. Therefore it is timely to re-examine the derivation of gene transport equations. Fisher's equation, which appeared in his seminal paper [8], is

$$\frac{\partial p}{\partial t} = k \frac{\partial^2 p}{\partial x^2} + mp(1 - p), \quad (1.1)$$

where p is the "frequency of the mutant gene" and m is "intensity of selection in favour of the mutant gene".

Fisher considered a population "distributed in a linear habitat . . . which it occupies with uniform density".

¹School of Mathematics and Applied Statistics, University of Wollongong, Wollongong NSW 2522, Australia; e-mail: pbroad@uow.edu.au.

²AgResearch Ltd., Wallaceville Animal Research Centre, PO Box 40063, Upper Hutt, New Zealand.

³School of Mathematics and Statistics, Australian Defence Force Academy, Canberra ACT 2600, Australia.

© Australian Mathematical Society 2002, Serial-fee code 1446-1811/02

The case of two possible pre-existing alleles occupying the locus of the mutant gene was simplified to that of a "... parent allelomorph, which we shall suppose to be the only allelomorph present". The physiological or anatomical characteristic governed by the mutant gene was assumed to be recessive, in keeping with the "common recessiveness of observed mutations".

In Section 2, we impose Fisher's assumptions on a spatially homogeneous population with discrete breeding cycles. This does not lead to the simple logistic reproduction rate $mp(1-p)$ but to another nonlinear growth rate for which the leading term is of the form $\kappa p^2(1-p)$, with κ constant. The cubic production term leads to a much delayed spread of the new gene after the mutation occurs. However for a species reproducing asexually, we do indeed recover the quadratic logistic reproduction law.

In Section 3, we re-examine the direct continuum modelling approaches of [2] and [13], leading to the same class of possible reaction-diffusion-convection equations for $p(x, t)$. We point out that in the common case of uniform selective advantage of the double recessive genotype over all others, the Huxley equation is appropriate. This is one of a small class of nonlinear reaction-diffusion equations that admits particular exact solutions by the Painlevé methods [5, 11] or by nonclassical symmetry reductions [3, 7]. The direct continuum modelling approach already shows us how to incorporate a non-uniform total population density. We construct a nontrivial exact solution for one such case.

2. Growth rate of a mutant population from discrete breeding cycles

At a single locus, there occurs one pair of genes, each having two possible allelomorphs, labelled 'a' (recessive and advantageous) and 'A' (dominant). In the classical Mendel binary scheme, the dominant physical characteristic associated with allele A will be equally apparent in the hybrid population with genotype (Aa) and in the pure-bred population with genotype (AA). Assume that these two genotypes have the same survival rate r_0 from the zygote to sexual maturity. The recessive characteristic 'a' is apparent only in the population with double recessive genotype (aa). This genotype will have a slightly higher survival rate r_2 . We assume random mating, with fertility independent of genotype. Following the notation of Fulford *et al.* [9, pp. 166–187], in the k th generation at conception, N_k^* is the total population, $N_k^*(AA)$, $N_k^*(Aa)$ and $N_k^*(aa)$ are the populations of the three genotypes, while corresponding symbols without asterisks refer to genotype populations in the k th generation at sexual maturity. Here $G_k(AA)$, $G_k(Aa)$ and $G_k(aa)$ are the genotype frequencies in the k th adult generation, for example $G_k(Aa) = N_k(Aa)/N_k$. Also $P_k(A)$ and $P_k(a)$ are the gene frequencies in the gene pool of the k th generation of adults.

By definition,

$$N_{k+1}^*(AA) = G_{k+1}^*(AA)N_{k+1}^*. \quad (2.1)$$

By the assumption of random mating, $G_{k+1}^*(AA)$ has an expected value of $[P_k(A)]^2$. For a large population, (2.1) reduces to

$$N_{k+1}^*(AA) = [P_k(A)]^2 N_{k+1}^*.$$

Similarly, we have

$$N_{k+1}^*(aa) = [P_k(a)]^2 N_{k+1}^* \quad \text{and} \quad N_{k+1}^*(Aa) = 2P_k(A)P_k(a)N_{k+1}^*.$$

By definition,

$$P_{k+1}(a) = \frac{2N_{k+1}(aa) + N_{k+1}(Aa)}{2N_{k+1}(aa) + 2N_{k+1}(Aa) + 2N_{k+1}(AA)}. \quad (2.2)$$

Now, in (2.2) we substitute

$$\begin{aligned} N_{k+1}(AA) &= r_0 N_{k+1}^*(AA), & N_{k+1}(Aa) &= r_0 N_{k+1}^*(Aa), \\ N_{k+1}(aa) &= r_2 N_{k+1}^*(aa) & \text{and} & \quad P_k(A) = 1 - P_k(a) \end{aligned}$$

to obtain

$$P_{k+1} = \frac{(\beta - 1)P_k^2 + P_k}{1 + (\beta - 1)P_k^2}, \quad (2.3)$$

where β is the relative fitness, $\beta = r_2/r_0$. Wherever the symbol P_k appears, it will be an abbreviation for $P_k(a)$. In the text [9, pp. 166–197], (2.3) was not compared to the commonly used polynomial growth laws. Now we simply note that in known beneficial mutations, $\beta - 1$ is usually small. Then (2.3) implies

$$P_{k+1} - P_k = (\beta - 1)P_k^2(1 - P_k) + O([\beta - 1]^2). \quad (2.4)$$

Note that the quadratic logistic growth term of Fisher's equation is replaced by cubic logistic growth. The above analysis is generalised here to the case of all three genotypes having different survival rates r_0 , r_1 and r_2 , with subscripts referring to the number of 'a' genes present. We obtain

$$\begin{aligned} P_{k+1} &= \frac{(\beta_2 - \beta_1)P_k^2 + \beta_1 P_k}{1 + 2(\beta_1 - 1)P_k + (1 + \beta_2 - 2\beta_1)P_k^2} \\ &= P_k + P_k(1 - P_k)([\beta_2 - \beta_1]P_k \\ &\quad + [\beta_1 - 1][1 - P_k]) + O([\beta_2 - 1]^2 + [\beta_1 - 1]^2). \end{aligned}$$

Now consider an *asexual* species that has a strain with phenotype A and another with slightly advantageous mutant phenotype a . Using a similar notation as before, we have

$$P_{k+1} = \frac{P_k N_{k+1}^*}{P_k N_{k+1}^* + \beta(1 - P_k)N_{k+1}^*},$$

implying

$$P_{k+1} - P_k = (\beta - 1)P_k(1 - P_k) + O([\beta - 1]^2).$$

The quadratic logistic source term of Fisher's equation is more appropriate for asexual species.

If Δt is the time between successive generations, then (2.4) is written as

$$\frac{P_{k+1} - P_k}{\Delta t} = \kappa P_k^2(1 - P_k),$$

where $\kappa = (\beta - 1)/\Delta t$. Assuming Fick's law with equal mobility for each genotype population, we obtain the Huxley equation

$$\frac{\partial p}{\partial t} = D \frac{\partial^2 p}{\partial x^2} + \kappa p^2(1 - p), \quad (2.5)$$

which should henceforth take more prominence in Mendelian genetics. In the next section, similar conclusions are drawn from direct continuum modelling.

3. Direct continuum models for genotype population densities

We consider one-dimensional population densities $\rho_0(x, t)$, $\rho_1(x, t)$ and $\rho_2(x, t)$ for genotypes with zero, one and two copies respectively of the recessive gene a . In this notation, Skellam's equations [13] are

$$\begin{aligned} \frac{\partial \rho_0}{\partial t} &= D \frac{\partial^2 \rho_0}{\partial x^2} - \mu \rho_0 + \gamma_0(1 - p)^2 \rho, \\ \frac{\partial \rho_1}{\partial t} &= D \frac{\partial^2 \rho_1}{\partial x^2} - \mu \rho_1 + \gamma_1 2p(1 - p) \rho, \\ \frac{\partial \rho_2}{\partial t} &= D \frac{\partial^2 \rho_2}{\partial x^2} - \mu \rho_2 + \gamma_2 p^2 \rho, \end{aligned}$$

where ρ is the total population density, $\rho = \rho_0 + \rho_1 + \rho_2$ and p is the frequency of the a -allele in the gene pool, $p = (2\rho_2 + \rho_1)/2\rho$. Here reproductive success rates γ_i may differ among the three genotype combinations $i = 0, 1, 2$. These three equations imply

$$\frac{\partial p}{\partial t} = D \left[\frac{\partial^2 p}{\partial x^2} + 2\eta \frac{\partial p}{\partial x} \right] + p(1 - p) [s_2 p + s_1(1 - p)], \quad (3.1)$$

where $s_1 = \gamma_1 - \gamma_0$, $s_2 = \gamma_2 - \gamma_1$ and

$$\eta(x, t) = \frac{\partial}{\partial x} \log \rho(x, t).$$

We could equally well have chosen a single common value γ for the reproductive success rates γ_i , and then modelled selective advantage by having different death rates μ_i for the three genotypes. However, as seen from (2.5) of the discrete model, this approach leads to the same source terms. In conjunction with direct continuum modelling [2] it leads also to the same convection and diffusion terms. In the case of constant total population density, $\eta = 0$, originally considered by Fisher, the three coupled equations for genotype densities imply a relatively simple single uncoupled reaction-diffusion equation for the gene frequency. However we point out here that for the case of principal interest, the Mendelian case wherein individuals with either one or two copies of the dominant A-gene have the same characteristics, $s_1 = 0$ and (3.1) reduces to Huxley's equation (2.5) with $\kappa = s_2$. Skellam [13] proceeded to analyse the effect of the extra convective term $2D\eta(x, t)\partial p/\partial x$ in (3.1). For this purpose, he chose $s_1 = s_2$ which resulted in a quadratic logistic source term. This choice afforded a simpler analysis since the comparative case with $\eta = 0$ then happened to be Fisher's equation (1.1), for which travelling wave solutions were already well known. However, there is no reason why this case should have any particular practical relevance. In the general case with $s_1 \neq 0$ and $s_2 \neq s - 1$, (3.1) is the Fitzhugh-Nagumo equation with an additional convective term, due in this case to the migratory diffusive flux of the total population.

In the case of two similar cohabiting strains of an asexual species distinguished by alternative characteristics A and a with densities ρ_0 and ρ_1 respectively, Skellam's reasoning [2] leads to

$$\frac{\partial p}{\partial t} = D \left[\frac{\partial^2 p}{\partial x^2} + 2\eta \frac{\partial p}{\partial x} \right] + sp(1 - p),$$

where $s = \gamma_1 - \gamma_0$ and

$$\eta = \frac{\partial}{\partial x} \log \rho, \quad \text{with} \quad \rho = \rho_1 + \rho_0.$$

Notice that in this case the product source term $sp(1 - p)$ is not due to heterosexual coupling.

In order to compare the effects of the Fisher quadratic a-allele source term $Q(p) = mp(1 - p)$ and the Huxley cubic source term $Q(p) = \kappa p^2(1 - p)$, we take $\kappa = 27m/16$, so that both source functions have the same maximum value $m/4$. It is straightforward to obtain the simple analytic solutions for the spatially uniform population $p(t)$. With initial condition $p(0) = 0.01$, there is a large time lag of the Huxley growth compared to the Fisher growth. The maximum growth rate under Huxley dynamics, appropriate for sexual species, takes twenty times longer to arrive than that of the Fisher dynamics, appropriate for asexual species. Although asexual species lack the genetic variability effected by the shuffling processes of meiosis and fertilisation, variations due to favourable mutations are likely to show up more quickly in the asexual population.

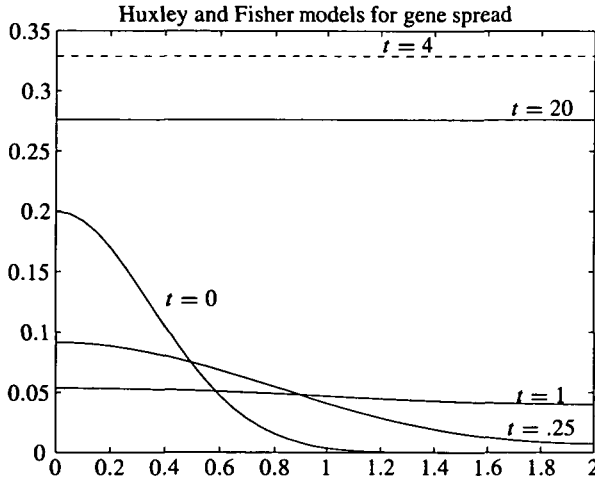


FIGURE 1. Spread of a Gaussian clump initially $0.2 \exp(-4x^2)$ according to (a) Huxley's equation with $D = 1$ and $\kappa = 1$ (solid curves) and (b) Fisher's equation (dashed curve) with the same maximum production rate.

Mutations in the sexual species, which are typically more complex, take longer to gain prevalence in the population. Since the spatially uniform solutions are stable, similar conclusions can be drawn for a spatially variable population mixture. In Figure 1, we display solutions of the two models with the same initial localized Gaussian clump of *a*-alleles, contained within the region $0 \leq x \leq 2$ by zero-flux boundary conditions $p_x = 0$.

This is a numerical method-of-lines solution obtained by using the program PDETWO of Melgaard and Sincovec [12]. At early times, the peak value of mutant frequency decreases as the mutant population spreads by diffusion. Here time scales and length scales have been chosen so that the diffusion coefficient and the growth coefficient are of order one. At a dimensionless time of the order of 1, the peak value reaches its minimum as the growth term begins to dominate the diffusion term. Thereafter, the solution resembles the spatially uniform solution. Mutant takeover is greatly retarded in the Huxley model compared to the Fisher model.

4. Nonclassical symmetry analysis of reaction-diffusion-convection equations

The most fundamental evolution equations of population genetics belong to the class of (1+1)-dimensional quasilinear reaction-diffusion-convection equations

$$0 = F(x, t, p, p_x, p_{xx}) = p_t - p_{xx} - 2\eta(x, t)p_x - Q(p). \quad (4.1)$$

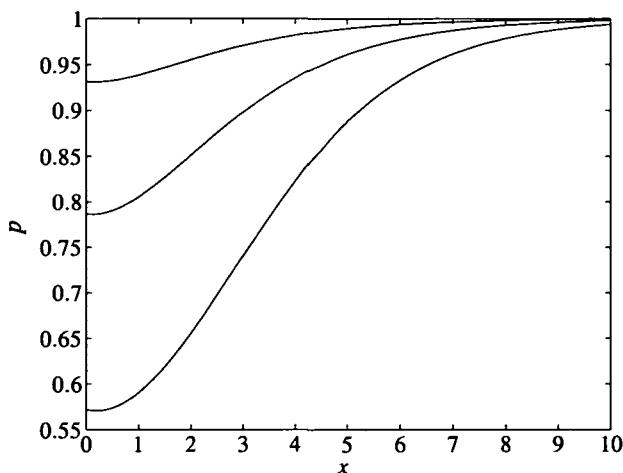


FIGURE 2. Solution at times $t = 0, 1$ and 2 to the Huxley equation with a convection term, subject to zero gradient at $x = 0$ and $x = 2$.

The most relevant cases, those of the asexually reproducing advantaged mutant, and the sexually reproducing recessive advantaged mutant, have a quadratic Fisher-type logistic growth function $Q = p(1 - p)$ and a cubic Huxley-type logistic growth function $Q = p^2(1 - p)$ respectively.

In a study of any practical class of nonlinear partial differential equations, the longest established systematic approach to finding exact solutions is the Lie symmetry classification. The symmetry group possessed by the generic reaction-diffusion equation [10] with $\eta = 0$ is nothing more than the translation group, generated by

$$\Gamma = a_1 \frac{\partial}{\partial t} + a_2 \frac{\partial}{\partial x},$$

with a_1 and a_2 constant. The group-invariant solutions are the steady states ($a_2 = 0$), the spatially uniform solutions ($a_1 = 0$) and the travelling wave solutions ($a_1 \neq 0$ and $a_2 \neq 0$). The travelling wave solution satisfies a nonlinear second-order ordinary differential equation for $p(s)$, where s is the d'Alembert variable $s = x - a_2 t/a_1$. Except for some special cases of wave speed [1], even the travelling wave solutions are difficult to obtain exactly. As a rule, symmetry classification of PDEs with free coefficient functions, such as $Q(p)$ or $\eta(x, t)$, results in special cases that are linear functions or products of simple powers, logarithms and exponentials of linear functions of the arguments. This is the case for reaction-diffusion equations [10]. Neither the Fisher equation nor the Huxley equation, whose polynomial source terms $Q(p)$ are more complicated than a simple power law, possesses additional classical symmetries.

We now investigate the possibility of *nonclassical* symmetries. Unlike the classical method, the nonclassical symmetry method [4] leads to *nonlinear* determining rela-

tions. However, following the advent of Clarkson and Kruskal's direct method [6], a number of new nonclassical solutions were constructed for practical nonlinear PDEs. Group-invariant solutions must satisfy the invariant surface condition (ISC)

$$Tp_t + Xp_x = P.$$

If we demand invariance of $F = 0$ subject to the constraint of the ISC, then this can sometimes lead to additional reductions to an ODE that are not obtainable by the classical method. In terms of the second prolongation $\Gamma^{(2)}$, the nonclassical symmetry determining relations are

$$\Gamma^{(2)}F = 0|_{F=0, \text{ISC}}. \quad (4.2)$$

After using the ISC to eliminate p_t and its derivatives, then subsequently using the governing equation to eliminate p_{xx} and higher order derivatives, (4.2) reduces to a polynomial equation in p_x , from which we obtain the determining relations for the coefficients $X(x, t, p)$, $T(x, t, p)$ and $P(x, t, p)$ of the infinitesimal symmetry generating operator Γ . In the case of the gene propagation equation (3.1) with Huxley source term $\kappa p^2(1 - p)$, we deduce that the nonclassical symmetry determining equations have a nontrivial solution if and only if η is a function of t alone. That is, the total population density must take the form $\rho = \rho_0(t)e^{\eta(t)x}$, which allows a restricted form of monotonic spatial variability in total population density at each time. In this special case, the convection term in the gene transport equation may be transformed to zero by a change of accelerating reference frame:

$$\begin{aligned} \bar{x} &= x + 2 \int_0^t D\eta(s) ds; & \bar{t} &= t \\ \frac{\partial p}{\partial \bar{t}} &= D \frac{\partial^2 p}{\partial \bar{x}^2} + \kappa p^2(1 - p). \end{aligned}$$

Remarkably, only this Huxley source term, and some other cubic source terms, result in genuine nonclassical symmetries [3, 7]. Further, the nonclassical reductions lead to an explicit symmetric solution [3, 5, 7, 11]. For $\kappa = 1$, this is

$$p = \frac{e^{\bar{t}/2 + \bar{x}/\sqrt{2}} + \sqrt{2}}{e^{\bar{t}/2 + \bar{x}/\sqrt{2}} + \bar{x} - \bar{t}\sqrt{2} + 3}. \quad (4.3)$$

This solution (4.3) has a local minimum at $\bar{x} = a(\bar{t})$, where

$$a(\bar{t}) = \sqrt{2} W \left(\sqrt{2} \exp \left[\frac{3}{\sqrt{2}} - 2 - \frac{3}{2} \bar{t} \right] \right) - 3 + 2\sqrt{2} + \bar{t}\sqrt{2},$$

where W is Lambert's function. We may now impose a familiar Neumann zero-gradient boundary condition by choosing

$$\bar{t} = t; \quad \bar{x} = x + a(t).$$

In terms of coordinates x and t , the governing gene transport equation is (4.1) with the particular choice $\eta = a'(t)/2D$. The solution with zero-gradient boundary condition is displayed in Figure 2. A population which has been elsewhere taken over by the a -allele, is initially depleted of a -alleles near the origin, for example by selective removal by pest controllers. The advantageous mutation then diffuses from the right to again take over the population near the origin.

Acknowledgements

We thank Nikolai Kudryashov, Clara Nucci, Mark Ablowitz and Angus Hurst who made helpful suggestions. We are grateful to Martin Kruskal and to the organizers of *Kruskal 2000*.

References

- [1] M. J. Ablowitz and A. Zeppetella, "Explicit solutions of Fisher's equation for a special wave speed", *Bull. Math. Biol.* **41** (1979) 835–840.
- [2] D. G. Aronson and H. F. Weinberger, "Nonlinear diffusion in population genetics, combustion, and nerve pulse propagation", in *Partial Differential Equations and Related Topics* (ed. J. A. Goldstein), Lect. Notes. in Math. 446, (Springer, Berlin, 1975) 5–49.
- [3] D. J. Arrigo, J. M. Hill and P. Broadbridge, "Nonclassical symmetry reductions of the linear diffusion equation with a nonlinear source", *IMA J. Appl. Math.* **52** (1994) 1–24.
- [4] G. W. Bluman and J. D. Cole, "The general similarity solution of the heat equation", *J. Math. Mech.* **18** (1969) 1025–42.
- [5] Z. Chen and B. Guo, "Analytic solutions of the Nagumo equation", *IMA J. Appl. Math.* **48** (1992) 107–115.
- [6] P. A. Clarkson and M. D. Kruskal, "New similarity reductions of the Boussinesq equation", *J. Math. Phys.* **30** (1989) 2201–13.
- [7] P. A. Clarkson and E. L. Mansfield, "Symmetry reductions and exact solutions of a class of nonlinear heat equations", *Physica D* **70** (1994) 250–288.
- [8] R. A. Fisher, "The wave of advance of advantageous genes", *Ann. Eugenics* **7** (1937) 335–369.
- [9] G. Fulford, P. Forrester and A. Jones, *Modelling with Differential and Difference Equations* (Cambridge University Press, New York, 1997).
- [10] V. A. Galaktionov, V. A. Dorodnitsyn, G. G. Elenin, S. P. Kurdyumov and A. A. Samarskii, "A quasilinear heat equation with a source: peaking, localization, symmetry, exact solutions, asymptotics, structures", *J. Soviet. Math.* **41** (1988) 1222–92.
- [11] N. A. Kudryashov, "Partial differential equations with solutions having movable first-order singularities", *Phys. Lett. A* **169** (1992) 237–242.

- [12] D. K. Melgaard and R. K. Sincovec, "General software for two-dimensional non-linear partial differential equations", *ACM Trans. Math. Software* 7 (1981) 106–125.
- [13] J. G. Skellam, "The formulation and interpretation of mathematical models of diffusionary processes in population biology", in *The Mathematical Theory of the Dynamics of Biological Populations* (eds. M. S. Bartlett and R. W. Hiorns), (Academic Press, London, 1973) 63–86.