Theoretical Population Biology xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Theoretical Population Biology

journal homepage: www.elsevier.com/locate/tpb

The infinitesimal model: Definition, derivation, and implications

N.H. Barton [a], A.M. Etheridge [b], A. Véber [c,*]

[a] Institute of Science and Technology, Am Campus I, A-3400 Klosterneuberg, Austria
[b] Department of Statistics, University of Oxford, 24–29 St Giles, Oxford OX1 3LB, UK
[c] Centre de Mathématiques Appliquées, École Polytechnique, Route de Saclay, 91128 Palaiseau Cedex, France

ARTICLE INFO

Article history:
Received 18 May 2017
Available online xxxx

Keywords:
Infinitesimal model
Selection
Epistasis
Quantitative genetics

ABSTRACT

Our focus here is on the *infinitesimal model*. In this model, one or several quantitative traits are described as the sum of a genetic and a non-genetic component, the first being distributed within families as a normal random variable centred at the average of the parental genetic components, and with a variance independent of the parental traits. Thus, the variance that segregates within families is not perturbed by selection, and can be predicted from the variance components. This does not necessarily imply that the trait distribution across the whole population should be Gaussian, and indeed selection or population structure may have a substantial effect on the overall trait distribution. One of our main aims is to identify some general conditions on the allelic effects for the infinitesimal model to be accurate. We first review the long history of the infinitesimal model in quantitative genetics. Then we formulate the model at the phenotypic level in terms of individual trait values and relationships between individuals, but including different evolutionary processes: genetic drift, recombination, selection, mutation, population structure, ... We give a range of examples of its application to evolutionary questions related to stabilising selection, assortative mating, effective population size and response to selection, habitat preference and speciation. We provide a mathematical justification of the model as the limit as the number M of underlying loci tends to infinity of a model with Mendelian inheritance, mutation and environmental noise, when the genetic component of the trait is purely additive. We also show how the model generalises to include epistatic effects. We prove in particular that, within each family, the genetic components of the individual trait values in the current generation are indeed normally distributed with a variance independent of ancestral traits, up to an error of order $1/\sqrt{M}$. Simulations suggest that in some cases the convergence may be as fast as $1/M$.

1. Introduction

The infinitesimal model is a simple and robust model for the inheritance of quantitative traits, in which these are the sum of a genetic and a non-genetic (environmental) component, and the genetic component of offspring traits follows a normal distribution around the average of the parents; this distribution has a variance that is independent of the parental trait values, and, in a large outcrossing population, the variance remains constant despite selection. With inbreeding, the variance decreases in proportion to relatedness. Of course, selection may cause the distribution across the whole population to deviate from normality. The crucial point is that under the infinitesimal model, the distribution of genetic components *within families* remains normal, with variance that evolves in a way that is entirely determined by relatedness.

This model has its roots in the observations of Galton (1877, 1885, 1889), and their analysis by Pearson (1896, 1897). Fisher (1918) showed that trait values and their (co)variances can be broken down into components, and that the phenotypic observation of constant within-family variance is consistent with a large number of Mendelian factors, with additive effects. The limiting infinitesimal model can be extended to include all the main evolutionary processes: recombination, mutation, random sampling drift, migration and selection. The model is hardly new, yet there seems to be no agreement on what precisely is meant by the infinitesimal model, nor on the conditions under which it is expected to apply. Moreover, although it has long been central to practical breeding, where it forms the genetic basis for the *animal model*, it is relatively little used in evolutionary modelling (see Kruuk, 2004; Hill and Kirkpatrick, 2010 for a review).

This paper provides a summary of the model, together with a rigorous derivation, including control over its accuracy as an approximation. We show that its predictions about within-family variance can be accurate even with epistasis. The reason can be

* Corresponding author.
E-mail addresses: nick.barton@ist-austria.ac.at (N.H. Barton), amandine.veber@cmap.polytechnique.fr (A. Véber).

http://dx.doi.org/10.1016/j.tpb.2017.06.001

understood intuitively, as follows. The classical theory of quantitative genetics gives a remarkably general description of evolution, in which the covariance in the values of a trait across individuals is a linear combination of a set of variance components, with coefficients determined by the probability of identity of sets of genes. Selection rapidly changes the trait mean, at a rate proportional to the additive genetic variance. However, when the trait depends on large numbers of genes, each of which makes a small contribution, selection has a negligible effect on the variance contributed by any individual locus. At the individual level, conditioning on the trait value hardly alters the distribution of effects of any one gene, at least in the short term; therefore, this distribution can be assumed constant. Importantly, it is not that allele frequencies do not change under the infinitesimal model: allele frequencies may change substantially due to random drift, mutation and migration; the key assumption is that selection only slightly perturbs the neutral distribution at any particular locus (Fisher, 1918; Robertson, 1960; Kimura, 1983, Ch. 6).

Our results here incorporate not only selection, but also mutation, random drift, population structure and some forms of epistasis. Dominance is left to future work. The evolutionary forces at work are captured by the actual pedigree of the population. Indeed, selection and structure pick out a particular pedigree, biased according to the trait values and the possible interactions between individuals. Thus, by conditioning on this pedigree and on the trait values in all previous generations, we are able to capture arbitrary forms of selection and population structure. The distribution of traits within families in the population is a multivariate normal distribution in which covariance is determined entirely by the pedigree and is independent of ancestral trait values. If some part of the pedigree or ancestral traits is unknown, then averaging with respect to the expected ancestral distribution, this multivariate normality is preserved. For example, it follows directly that conditioning on knowing just some of the trait values in the pedigree shifts the mean trait values in other families by a linear function of the conditioned values, but leaves variances within families unaltered.

After outlining the history of the infinitesimal model, we define it directly as a model for the distribution of phenotypes in a population; such a formal definition seems to be new. Initially, we implicitly assume an additive trait, but include all the usual evolutionary processes. For simplicity, we neglect linkage throughout. Having explained the phenotypic model, not only defining it at the level of the individual, but also showing how it can be simulated at the level of the population, we outline some of its applications. We then show that we can derive this infinitesimal model as the limit of a model of Mendelian inheritance, showing the conditions under which it is accurate and obtaining explicit bounds on the error. Finally, we show how the infinitesimal model extends to allow for epistasis, before presenting simulations that illustrate the main results.

We emphasise that our derivation of the infinitesimal model is distinct from earlier work, which used multi-locus models to analyse the effects of selection on complex traits (e.g. Bürger, 2000; Turelli and Barton, 1994; Kirkpatrick et al., 2002). The aim there was to connect population with quantitative genetics, and specifically, to find ways to approximate the effects of selection on the genetic variance, given a finite number of loci. In particular, Turelli and Barton (1994) investigated whether the trait distribution *across the whole population* could be approximated by a normal distribution. In contrast, here we aim to show that in the infinitesimal limit, the trait distribution *within families* is normally distributed, with a variance that is determined by the variance in the ancestral population and the pedigree relating individuals in those families, without making any detailed assumptions about the genetic basis of the trait, or about the form of the distribution of the trait across the population. Thus, we aim at a radical simplification of quantitative genetics.

2. The classical model

2.1. History

Although the infinitesimal model is named for its justification as the limit of infinitely many Mendelian genes, it can be defined purely phenotypically, and its origins trace back well before the rediscovery of Mendel's work in 1900. Here, we summarise the origins of the infinitesimal model, after which we will formulate a precise definition at the phenotypic level, with no explicit genetic assumptions.

In one of the earliest quantitative discussions of heredity, Fleeming Jenkin (1867) argued that blending inheritance could have no effect in the long term: a white man stranded on an inhabited tropical island would leave offspring who, over successive generations, would approach ever closer to the dark-skinned native population. Davis (1871) pointed out that in a large and stable population, an individual is expected to leave two children, four grandchildren, and so on, so that his total expected contribution is constant through time. Nevertheless, if offspring are precisely intermediate between their parents, the range of variation in the population must necessarily decrease. Darwin saw this as a serious problem for his theory, which required a source of variation to counter blending inheritance. (See Bulmer, 2004, for a detailed discussion of Jenkin's argument.)

Francis Galton gathered extensive data on the inheritance of continuous traits, and introduced many ideas that are now central to quantitative genetics. In experiments with sweet peas, he showed that seeds of offspring grown from seeds of different weights followed a normal distribution with a mean that reverted towards the population mean, and with variance independent of the parents' weight: "I was certainly astonished to find the family variability of the produce of the little seeds to be equal to that of the big ones, but so it was, and I thankfully accept the fact, for if it had been otherwise, I cannot imagine, from theoretical considerations, how the problem could be solved" (Galton, 1877, p. 513). (In Galton's experiments with sweet peas, plants were self-fertilised, so that the variance in families is, in fact, expected to decrease.) He saw a similar pattern for human height, and showed that the joint distribution of offspring and mid-parent is bivariate normal (Galton, 1885). Moreover, he understood that the variance of the population could remain stable under the joint influence of random mating, reversion of offspring towards the population mean, and generation of variance amongst offspring. Galton (1877) calculated the equilibrium variance, allowing for Gaussian stabilising selection, a calculation next made by Bulmer (1971) and Cavalli-Sforza and Bodmer (1971), nearly a century later.

Galton (1885, 1889) tried to explain his observations by formulating his 'law of ancestral heredity', which divided an individual's phenotype into geometrically declining contributions from parents, grandparents, great-grandparents, …; he interpreted this contribution from distant ancestors as being due to inherited factors which have some probability, p, of being expressed in each generation. Bulmer (1998) shows that Galton's law is equivalent to the quantitative genetics of an additive trait, with p being replaced by the heritability, $h^2 = V_A/V_P$ (where V_P is the total phenotypic variance and V_A the additive genetic variance of the trait); however, h^2 may vary from trait to trait, whereas Galton assumed that it is a constant parameter of the mechanism of inheritance. Galton's model explains reversion of offspring towards the population mean as being due to expression of factors inherited from earlier generations (Lush, 1937, p. 47). In contrast, under Mendelian inheritance, reversion to the mean arises because selection acts on the phenotypic variance, V_P, whereas only additive genetic variation, V_A, is passed on; the deviation of offspring is therefore $h^2 = V_A/V_P$

Please cite this article in press as: Barton, N.H., et al., The infinitesimal model: Definition, derivation, and implications. Theoretical Population Biology (2017), http://dx.doi.org/10.1016/j.tpb.2017.06.001.