Theoretical Population Biology 118 (2017) 46–49

Contents lists available at ScienceDirect

Theoretical Population Biology

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

Commentary

Commentary: Fisher's infinitesimal model: A story for the ages

Michael Turelli

Department of Evolution and Ecology, University of California, Davis, CA, USA

ARTICLE INFO

Article history:
Received 18 September 2017
Available online 5 October 2017

Keywords:
Polygenic traits
Gaussian approximations
Selection response
Additive effects
Epistasis
Linkage disequilibria

ABSTRACT

Mendel (1866) suggested that if many heritable "factors" contribute to a trait, near-continuous variation could result. Fisher (1918) clarified the connection between Mendelian inheritance and continuous trait variation by assuming many loci, each with small effect, and by informally invoking the central limit theorem. Barton et al. (2017) rigorously analyze the approach to a multivariate Gaussian distribution of the genetic effects for descendants of parents who may be related. This commentary distinguishes three nested approximations, referred to as "infinitesimal genetics," "Gaussian descendants" and "Gaussian population," each plausibly called "the infinitesimal model." The first and most basic is Fisher's "infinitesimal" approximation of the underlying genetics – namely, many loci, each making a small contribution to the total variance. As Barton et al. (2017) show, in the limit as the number of loci increases (with enough additivity), the distribution of genotypic values for descendants approaches a multivariate Gaussian, whose variance–covariance structure depends only on the relatedness, not the phenotypes, of the parents (or whether their population experiences selection or other processes such as mutation and migration). Barton et al. (2017) call this rigorously defensible "Gaussian descendants" approximation "the infinitesimal model." However, it is widely assumed that Fisher's genetic assumptions yield another Gaussian approximation, in which the distribution of breeding values in a population follows a Gaussian — even if the population is subject to non-Gaussian selection. This third "Gaussian population" approximation, is also described as the "infinitesimal model." Unlike the "Gaussian descendants" approximation, this third approximation cannot be rigorously justified, except in a weak-selection limit, even for a purely additive model. Nevertheless, it underlies the two most widely used descriptions of selection-induced changes in trait means and genetic variances, the "breeder's equation" and the "Bulmer effect." Future generations may understand why the "infinitesimal model" provides such useful approximations in the face of epistasis, linkage, linkage disequilibrium and strong selection.

1. Introduction

Fisher's 1918 paper, "The correlation between relatives under the supposition of Mendelian inheritance", is a mathematical landmark. It reconciled Mendelian inheritance with the patterns of variation and covariation among relatives for continuously varying characters compiled by Galton (1889), and it helped resolve the debate between the Mendelians and the Biometricians over the heritable basis of evolutionary change (Provine, 1971). The path to theoretical reconciliation of Mendel and Galton through the works of Yule and Pearson was summarized in Lock (1906, Ch. VIII) and Provine (1971). East (1916) played a pivotal role in synthesizing the empirical evidence on which the reconciliation was based (Turelli, 2016). At the centenary of Fisher (1918), new mathematical and experimental approaches have rekindled interest in what has become known as "the infinitesimal model", which postulates that continuous trait variation is often produced by "many" segregating Mendelian loci, each making a "small" contribution to the total variance, plus environmental effects. Barton et al. (pp. 50–73) present the most recent mathematical development, illuminating genetic conditions under which (unselected) descendants of known parents have approximately Gaussian-distributed genotypic values, whose variance depends only on the relationship of the parents and not on their phenotypes.

The infinitesimal model, popularized by Fisher (1918), is a classic story that reconciles Galton's observations (summarized in Galton, 1889) concerning the distribution and inheritance of continuously distributed phenotypes, such as human height and the length of *Pisum sativum* seed pods, with Mendelian genetics. Like many classic stories, its origin and content are a bit vague and its implications subject to multiple interpretations (*e.g.*, Wright, 1952; Provine, 1971; Bulmer, 1980, pp. 10–15; Hill, 1984a, pp. 8–19; Hill, 2013, p. 4; Barton et al., pp. 50–73). The initial insight was Mendel's (1866) realization that if a trait is determined by contributions from many independently inherited, dichotomous "factors", an essentially continuous distribution of phenotypes

E-mail address: mturelli@ucdavis.edu.

https://doi.org/10.1016/j.tpb.2017.09.003
0040-5809/

emerges, with continuity enhanced by non-heritable environmental contributions. The canonical text is Fisher (1918), important to both geneticists and statisticians for partitioning phenotypic variance into components associated with alleles, one-locus genotypes, two-locus interactions and environmental effects. Fisher (1918) focused on understanding the patterns of correlations between the phenotypes of relatives, but the most influential applications of what has come to be known as "the infinitesimal model" (cf. Bulmer, 1971) have focused on estimating the components of genetic and environmental variance, often with the goal of understanding the consequences of artificial and natural selection. My commentary describes the model (sort of), its history, some applications, and recent developments, especially Barton et al. (pp. 50–73). I emphasize a distinction between the underlying genetic assumptions ("infinitesimal genetics"), as exemplified by Fisher's (1918) treatment, and two closely related, but distinct, mathematical approximations inspired by them. Hill (2013) provides a more detailed overview focused on animal breeding, reviewing both theory and data.

2. What is the infinitesimal model?

Fisher (1918) was characteristically vague: "The simplest hypothesis … is that such features as stature are determined by a large number of Mendelian factors …" He introduced two-locus epistasis and analyzed dominance, unequal allelic effects and arbitrary allele frequencies across loci. But when illustrating the approach to normality as the number of loci increases, Fisher (1918, p. 402) implicitly assumed linkage equilibrium, comparable and (mainly?) additive effects, and explicitly assumed that the contributions of individual loci are negligible relative to the total genetic variance. He trod lightly on "avoidable complications", like multilocus epistasis, linkage and linkage disequilibrium, but concluded: "In spite of this, it is believed that the statistical properties of any feature determined by a large number of Mendelian factors have been successfully elucidated" (Fisher, 1918 p. 432). Any questions?

3. What are the implications of the infinitesimal model?

Galton (1889), which inspired Fisher (1918), focused on four observations: (1) the approximately Gaussian distribution of phenotypes for many continuous traits, (2) patterns of phenotypic correlations among relatives, (3) the linearity of regression of offspring phenotypes on parents, and (4) the invariance of within-family variance across parental phenotypes (summarized in Roughgarden, 1979, Ch. 9). Fisher (1918) addressed the Gaussian distribution of genetic effects by assuming polygenic inheritance (Mather, 1943) and calculated genetic covariances produced by shared alleles and one-locus genotypes. In their careful mathematical treatment, Barton et al. (pp. 50–73) analyze conditions that produce a multivariate Gaussian distribution for the genotypic values of the descendants of known parents (at least before selection acts on the descendants), as the number of loci approaches infinity. I will term this rigorously supportable asymptotic approximation, which follows from Fisher's infinitesimal-genetics assumptions, the "Gaussian descendants" approximation.

However, many applications of "the infinitesimal model" focus on selection and use what might be termed the "Gaussian population" approximation. This approximation, which underlies the "breeders equation" and the "Bulmer effect" predictions for changes in the population mean and additive genetic variance, assumes (without formal justification) that if many loci contribute to a character, the distribution of breeding values in a population will be approximately Gaussian, as will the joint distribution of genotypic values and phenotypes for parents and offspring—even in populations subject to recurrent selection on the character. With bivariate normality for the genotypic values of parents and offspring, linear regression and constant within-family variances follow from the properties of the multivariate Gaussian elucidated by Pearson (1896, 1903).

4. How about selection?

A central part of the debate between the Mendelians and Biometricians concerned the efficacy of natural selection on continuous variation as opposed to discrete differences (Provine, 1971). Given that Fisher championed the predominant role of natural selection in evolution and must have known Pearson's (1896, 1903) work on selection and the multivariate normal, it seems odd that neither Fisher (1918) nor Fisher's subsequent book, which focused on natural selection (Fisher, 1930), quantified the effects of selection on continuous traits. If the phenotypes of offspring and midparents follow a bivariate normal, selection response follows

$$R = \beta_{O|M} S, \tag{4.1}$$

where S is the difference between the mean phenotypes of selected parents and the population from which they are drawn, R is the difference between the mean of offspring produced with selection and the mean of offspring that would be produced without selection, and $\beta_{O|M}$ is the regression of offspring phenotype on the parental mean (midparent) (Bulmer, 1980 Ch. 9). Assuming that parents and offspring experience the same range of environments and that only shared alleles (without epistasis) produce the phenotypic similarity between offspring and parents, the results of Fisher (1918) imply $\beta_{O|M} = h^2 = V_A/V_P$, where V_A is the additive genetic variance and V_P is the phenotypic variance. Thus (4.1) becomes the "breeders equation"

$$R = h^2 S. \tag{4.2}$$

Oddly, this fundamental equation appeared first as a footnote buried in Lush (1937, p. 84); see Lush (1935), for an explicit acknowledgment of Sewall Wright's help). In addition to the classical approximation (4.2), Lush (1937) noted that a more accurate approximation for the numerator of $\beta_{O|M}$ includes contributions from higher-order additive terms, V_{AA}, V_{AAA}, etc. Given the centrality of (4.2) in animal and plant breeding (after Lush, 1937) and evolutionary theory (after Lande, 1976), one expects more fanfare at its birth. Obviously Wright and Fisher had no idea how accurate (or useful) the approximation would be, but they must have understood that many assumptions were involved. As described in the commentaries of Hill (1984a, b, 2013), quantitative geneticists focused for decades on practical applications, experimental tests, and extensions of (4.2), rather than genetic or mathematical justifications of this approximation.

Bulmer (1971) first described the consequences of selection-induced linkage disequilibrium for changing the additive variance that enters (4.2). As the number of loci, n, increases, selection intensity per locus declines with n, but the $n(n-1)/2$ linkage disequilibria produced by selection have a cumulative effect that does not vanish. Thus, even though the "genic variance", defined as the additive variance achieved with all relevant loci in linkage equilibrium, remains constant in this "infinitesimal limit", the additive variance (which involves variances and covariances across loci) changes—in a way that is independent of genetic details, depending only on heritability and the selection-induced change in phenotypic variance. Bulmer (1971) derived his expression for the change in additive variance from a regression argument (based on the Gaussian-population approximation), but it also follows from explicit multilocus population genetics (Turelli and Barton, 1994). However, these multilocus calculations also demonstrate that most forms of selection generate multilocus associations

that preclude a Gaussian distribution of genotypic values, even without dominance or epistasis, no matter how many loci contribute (Turelli and Barton, 1990). Hence, unlike the Gaussian-descendants approximation considered by Barton et al. (pp. 50–73), the Gaussian-population approximation does not hold as an asymptotic limit with large numbers of loci unless selection is Gaussian or very weak. Surprisingly, the deviations from (4.2), which emerge even for weak selection, remain relatively small under random mating for a wide range of models with intense selection and only moderate numbers of loci (Turelli and Barton, 1994). This supports the standard regression approximations that underlie the predictions for how selection changes the population mean (4.2) and additive genetic variance (Bulmer, 1971).

Understanding the dynamics of genic variance associated with allele-frequency changes – and the maintenance of genic variance – depends critically on the underlying genetics (Barton and Turelli, 1987). Yet, for many evolutionary questions, such details are likely to be irrelevant. One can simply invoke the relative robustness of (4.2) to investigate phenotypic evolution assuming constant heritability (e.g., Lande, 1976; Price et al., 1993) or use a testable model for changes in additive variance (e.g., Ashander et al., 2016). Alternatively, one can empirically study polygenic adaptation and its genetic basis by associating small allele-frequency differences found in genome-wide association studies with phenotypic differences and/or environmental variables (Turchin et al., 2012; Berg and Coop, 2014; Berg et al., 2017).

5. The "infinitesimal model" and the "Gaussian descendants" approximation

Barton et al. (pp. 50–73) characterize the "infinitesimal model" in terms of the genetic variance within families and its independence of parental phenotypes. As they note, assuming a Gaussian distribution of breeding values within a family is central to the "animal model" approach to estimating quantitative genetic parameters (Lynch and Walsh, 1998). The Gaussian-descendants approximation is valuable because it provides a rigorous basis for understanding the consequences of selection, even when the Gaussian-population approximation becomes untenable. In the parallel universe of theoretical evolutionary ecology, Bossert (unpublished) pioneered a phenotypic version of this model, in which inheritance is represented by a fixed "segregation kernel" that approximates the distribution of full-sib phenotypes as a Gaussian with mean equal to the midparent value and fixed variance, independent of the midparent. Bossert's (1963) approach was adapted by Slatkin (1970), Roughgarden (1972) and Slatkin and Lande (1976) to address changes over time and equilibria for population means and variances under various scenarios, including frequency- and density-dependent selection, that may produce non-Gaussian distributions of phenotypes in the population. Similarly, Turelli and Barton (1994) assumed that the distribution of breeding values within families was Gaussian and used alternative approximations to understand the population distribution of breeding values that would emerge under truncation and disruptive selection.

Barton et al. (pp. 50–73) ask what genetic models and what evolutionary processes are consistent with genetic values being approximately Gaussian among descendants, with variance independent of parental phenotypes and determined only by the initial genic variance in the population and the relatedness of the parents (see their Eq. (9)). Assuming only additive effects and no linkage, they show that even when mutation, selection, drift, population structure and gene flow act, the Gaussian-descendants approximation can be justified (which they call "the infinitesimal model", see their Eq. (13) for a rigorous treatment of mutation). Using modern versions of the central limit theorem that allow for non-identically distributed random variables and some departures from independence, they prove asymptotic convergence of the joint distribution of genetic effects for a set of descendants to a multivariate Gaussian, whose variance–covariance matrix depends only on the relationships of the parents (see their Eqs. (9)–(11)). Moreover, they bound the departures from the Gaussian as a function of the number of loci (see their Eqs. (10) and (15)) and show how some forms of epistasis can be accommodated (see their Section 3.2). It is important to note that their result applies to the joint distribution of descendants (who have not experienced selection), even when parents are selected, because Mendelian segregation and assortment produce the mixing needed to achieve normality for their progeny (even allowing for some linkage). However, the bivariate distribution of *parents* and offspring will not be Gaussian, unless Gaussian selection is applied to the parents.

6. Coda

The literature includes at least three interpretations of "the infinitesimal model": (1) Fisher's invocation of "…a large number of Mendelian factors…", each making a "small" contribution to the total genetic variance; (2) the Gaussian-descendants approximation, rigorously addressed by Barton et al. (2017, pp. 50–73), which can be supported asymptotically (as the number of loci increases) even when selection acts; and (3) the Gaussian-population approximation, widely used by both breeders and evolutionists to understand the consequences of selection, even though it can be rigorously justified only under Gaussian (or very weak) selection. Fisher's genetic assumptions and the Gaussian approximations all find support in new empirical analyses suggesting that hundreds of loci contribute to variation of traits, like human height, that first attracted Galton's attention (Lango Allen et al., 2010; Marouli et al., 2017). However, even with an effectively infinite number of loci contributing to variation, it remains a challenge to understand why Gaussian approximations are useful in the face of the extensive non-linearity (especially multilocus epistasis) we expect in genotype–phenotype maps (e.g., Hill et al., 2008; Barton et al., pp. 50–73) and to provide empirically useful bounds on the accuracy of the alternative Gaussian approximations. Barton et al. (pp. 50–73) set a new benchmark for understanding the asymptotic Gaussian-descendants approximation. Nevertheless, fully understanding "the infinitesimal model" will surely occupy many future generations.

Acknowledgments

I thank N.H. Barton, W.H. Bossert, M.G. Bulmer, G. Coop, W.G. Hill, C.H. Langley, N.A. Rosenberg, M. Slatkin and J.B. Walsh for instructive discussions and constructive suggestions that greatly clarified my presentation.

References

Ashander, J., Chevin, L.-M., Baskett, M.L., 2016. Predicting evolutionary rescue via evolving plasticity in stochastic environments. Proc. R. Soc. B 283, 20161690.

Barton, N.H., Etheridge, A.M., Veber, A., 2017. The infinitesimal model: definition, derivation, and implications. Theor. Popul. Biol. 118, 50–73.

Barton, N.H., Turelli, M., 1987. Adaptive landscapes, genetic distance and the evolution of quantitative characters. Genet. Res. 49, 157–173.

Berg, J.J., Coop, G., 2014. A population genetic signal of polygenic adaptation. PLOS Genet. 10, e1004412.

Berg, J.J., Zhang, X., Coop, G., 2017. Polygenic adaptation has impacted multiple anthropometric traits. BIORXIV-2017-167551v1.

Bossert, W.H., 1963. Simulation of Character Displacement. Harvard University, Cambridge MA, (unpublished).

Bulmer, M.G., 1971. The effect of selection on genetic variability. Am. Nat. 105, 201–211.

Bulmer, M.G., 1980. The Mathematical Theory of Quantitative Genetics. Clarendon Press, Oxford.
East, E.M., 1916. Studies of size inheritance in Nicotiana. Genetics 1, 164–176.
Fisher, R.A., 1918. The correlation between relatives under the supposition of Mendelian inheritance. Trans. Roy. Soc. Edinburgh 52, 399–433.
Fisher, R.A., 1930. The Genetical Theory of Natural Selection. Clarendon Press, Oxford.
Galton, F., 1889. Natural Inheritance. MacMillan, London.
Hill, W.G., 1984a. Quantitative Genetics, Part I. Explanation and Analysis of Continuous Variation. Van Nostrand Reinhold, New York.
Hill, W.G., 1984b. Quantitative Genetics, Part II. Selection. Van Nostrand Reinhold, New York.
Hill, W.G., 2013. Applications of population genetics to animal breeding, from wright, fisher and lush to genomic prediction. Genetics 196, 1–16.
Hill, W.G., Goddard, M.E., Visscher, P.M., 2008. Data and theory point to mainly additive genetic variance for complex traits. PLOS Genet. 4, e1000008.
Lande, R., 1976. Natural selection and random genetic drift in phenotypic evolution. Evolution 30, 314–334.
Lango Allen, H., Estrada, K., Lettre, G., et al., 2010. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467, 832–838.
Lock, R.H., 1906. Recent Progress in the Study of Variation, Heredity, and Evolution. John Murray, London.
Lush, J.L., 1935. Progeny test and individual performance as indicators of an animal's breeding value. J. Dairy Sci. 18, 1–19.
Lush, J.L., 1937. Animal Breeding Plans. Collegiate Press, Ames, Iowa.
Lynch, M., Walsh, B., 1998. Genetics and Analysis of Quantitative Traits. Sinauer, Sunderland, MA.
Marouli, E., Graff, M., Medina-Gomez, C., et al., 2017. Rare and low-frequency coding variants alter human adult height. Nature 542, 186–190.
Mather, K., 1943. Polygenic inheritance and natural selection. Biol. Rev. 18, 32–64.
Mendel, G., 1866. Verusche ueber plaflzenhybriden. Verhandlungen des naturforschenden Vereines in Brunn, Bd. IV fur das Jahr 1865:3–47 (English translation available at: http://www.esp.org/foundations/genetics/classical/gm-65.pdf).
Pearson, K., 1896. Mathematical contributions to the theory of evolution. III. regression, heredity and panmixia. Philos. Trans. R. Soc. Lond. Ser. A 187, 253–318.
Pearson, K., 1903. Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variability and correlation of organs. Philos. Trans. R. Soc. Lond. Ser. A 200, 1–66.
Price, T., Turelli, M., Slatkin, M., 1993. Peak shifts produced by correlated response to selection. Evolution 47, 280–290.
Provine, W.B., 1971. The Origins of Theoretical Population Genetics. University of Chicago Press, Chicago, IL.
Roughgarden, J., 1972. The evolution of niche width. Am. Nat. 106, 683–718.
Roughgarden, J., 1979. Theory of Population Genetics and Evolutionary Ecology: an Introduction. MacMillan, New York.
Slatkin, M., 1970. Selection and polygenic characters. Proc. Natl. Acad. Sci. USA 66, 87–93.
Slatkin, M., Lande, R., 1976. Niche width in a fluctuating environment–density independent model. Am. Nat. 110, 31–55.
Turchin, M.C., Chiang, C.W.L., Palmer, C.D., Sankararaman, S., Reich, D., et al., 2012. Evidence for widespread selection on standing variation in Europe at height-associated SNPs. Nature Genet. 44, 1015–1019.
Turelli, M., 2016. Edward East on the Mendelian basis of quantitative trait variation. Genetics 204, 1321–1323.
Turelli, M., Barton, N.H., 1990. Dynamics of polygenic characters under selection. Theor. Popul. Biol. 38, 1–57.
Turelli, M., Barton, N.H., 1994. Genetic and statistical analyses of strong selection on polygenic traits: What, me normal? Genetics 138, 913–941.
Wright, S., 1952. The genetics of quantitative variability. In: Reeve, E.C.R., Waddington C.H. (Eds.), Quantitative Inheritance. Her Majesty's Stationary Office, London, pp. 5–41.