Studies in History and Philosophy of Biol & Biomed Sci 79 (2020) 101226

Contents lists available at ScienceDirect

Studies in History and Philosophy of Biol & Biomed Sci

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

The origins of the stochastic theory of population genetics: The Wright-Fisher model

Yoichi Ishida[a,*], Alirio Rosales[b]

[a] Department of Philosophy, Ohio University, Ellis Hall 201, Athens, OH 45701, USA
[b] Department of Zoology annd Biodiversity Research Centre, University of British Columbia, 2212 Main Mall, Vancouver, BC V6T 1Z4, Canada

1. Introduction

In the stochastic theory of population genetics, the Wright-Fisher model of genetic drift is historically the most influential model for studying random fluctuations in the genetic variation of a population.[1] The model is canonically formulated as follows: A population of diploid organisms has constant size N so that there will be $2N$ genes at a chromosomal locus, and assume that genes in a locus of our interest come in one of two allelic forms, A or a. Apart from finite population size, there are no other evolutionary factors, such as selection, mutation, or migration.[2] Given these assumptions and others to be discussed below, the model describes random fluctuations in the number of A alleles, which ranges from 0 to $2N$. Let p_{ij} denote the probability that the number of A alleles changes from i to j. The canonical formulation of the Wright-Fisher model then says:

$$p_{ij} = \binom{2N}{j}\left(\frac{i}{2N}\right)^{j}\left(1 - \frac{i}{2N}\right)^{2N-j} \tag{1}$$

where the first term on the right-hand side is the binomial coefficient and the entire right-hand side defines a binomial distribution. Here p_{ij} is called the transition probability of a Markov chain, a type of stochastic process where the probability distribution of future states of a system depends on its present state but not on any of its past states (Ewens, 2004, pp. 20–21; Hartl & Clark, 2007, p. 102). Equation (1) thus specifies the probability that the system moves from state i to a possible future state j.[3]

The canonical formulation of the Wright-Fisher model raises important questions for understanding the history of theoretical population genetics. The model is attributed to the work done in the 1920s by R. A. Fisher (1922a, 1930) and Sewall Wright (1931), but a systematic theory of Markov chains or that of stochastic processes was not available at that time. The general theory of stochastic processes began with A. N. Kolmogorov's 1931 paper (Kolmogorov, 1931; Dynkin, 1989), and the theory of Markov chains became an active area of mathematical research only in the 1930s (see Section 3).[4] Moreover, it was Gustave Malécot (1944) who introduced Equation (1) to population genetics (Seneta, 1974; Edwards, 1978; Nagylaki, 1989; Epperson, 1999). Nevertheless, Fisher and Wright did use a model of drift based on a binomial distribution, and their mathematical conclusions essentially agree with those derived later from the Markov-chain model (see, e.g., Ewens, 2004, pp. 20–31). What, then, was the model that Fisher and Wright actually used? How did their model differ from Malécot's? How did the Markov-chain formulation of the Wright-Fisher model become established among population geneticists? And what conceptual changes, if any, did the Markov-chain formulation bring about?[5]

To answer these questions, we begin by reconstructing Fisher's original model, which we call *the binomial model*. In the 1920s, Fisher developed his theory of evolution (Fisher, 1922a, 1930) along with his theory of mathematical statistics (Fisher, 1922b, 1925a, b). This par-

---

* Corresponding author.
E-mail addresses: ishiday@ohio.edu (Y. Ishida), arosales@zoology.ubc.ca (A. Rosales).

[1] Neuhauser (2001) and Ewens (2004, Ch. 3) present the basic mathematical models of the stochastic theory of population genetics.

[2] In other words, this is "the 'simple' Wright-Fisher model" (Ewens, 2004, p. 21), and more complicated Wright-Fisher models incorporate other evolutionary factors (see, e.g., Ewens, 2004, p. 21).

[3] We discuss Markov chains in more detail in Section 3.

[4] An English translation of Kolmogorov (1931) is available in Shiryayev (1992, pp. 62–108).

[5] For the background history of population genetics, see Provine (1971, 1978, 1985, 1986, 1990); Beatty (1987, 1992); Turner (1987); Hodge (1992, 2011); Edwards (2001); Skipper (2002); Plutynski (2005, 2007); Ishida (2009, 2017); Dietrich and Skipper (2013); and Rosales (2017). See also Adams (1968, 1970); Burian and Gayon (1999); Charlesworth and Charlesworth (2017); and Dung et al. (2019). Brief biographies, chronologies, and discussions of Malécot's work can be found in Seneta (1974), Crow (1987), Nagylaki (1989), Epperson (1999), and Slatkin and Veuille (2002). These works do not provide detailed historical analyses but explain Malécot's work from a contemporary viewpoint.

https://doi.org/10.1016/j.shpsc.2019.101226
Received 12 May 2019; Received in revised form 30 August 2019; Accepted 6 October 2019

allel is often overlooked in the historical and philosophical studies of Fisher's evolutionary theory.[6] But we show in Section 2 that it is important for understanding how Fisher based the binomial model on his theory of mathematical statistics and how the statistical concepts of random sampling, sampling error, and sampling distributions became relevant to modeling in theoretical population genetics.

In Section 3, we describe how Malécot reformulated the binomial model as a Markov chain and discuss one of the two consequences of his reformulation. The theory of stochastic processes, which includes the theory of Markov chains, was conceived as a probabilistic analogue of a deterministic physical theory like Newtonian mechanics (e.g., Kolmogorov, 1931). Thus, one consequence of Malécot's reformulation was the theoretical treatment of an evolving population as a dynamical system governed by probabilistic laws, such as the one given by Equation (1).

Another consequence of Malécot's reformulation, to which we turn in Section 4, was the emergence of what we call *the Markovian path* of theoretical population genetics. In the 1950s, Malécot's work on the Markov-chain model was introduced to the English-speaking mathematicians and population geneticists by William Feller (1950, p. 315; 1951, pp. 231–232).[7] A small group of mathematicians and population geneticists then began to pursue the project of providing mathematically rigorous foundations for the theoretical study of drift initiated by Fisher. In the 1960s, English-language textbooks on theoretical population genetics began to appear (Moran, 1962; Ewens, 1969), and they discussed the Markov-chain model. In addition, several journals were founded between 1960 and 1970 that would publish mathematical papers on stochastic population genetics.[8] By the end of the 1960s, the Markov-chain formulation of Fisher's model became canonical, and the model so formulated was attributed to Fisher and Wright.

Our historical account shows that the origins of the stochastic theory of population genetics include Fisher's binomial model, which set the stage for Malécot's reformulation of the model as a Markov chain, and the emergence of the Markovian path, which adopted the theory of stochastic processes as a mathematical foundation of theoretical population genetics. In Section 5, we briefly discuss how this history illuminates philosophical research on drift.

2. Fisher's binomial model

In 1921, Hagedoorn and Hagedoorn (1921) suggested that random extinction of genes from a population is a more important influence than natural selection on the evolution of species. To evaluate this claim, Fisher (1922a) developed the binomial model, considering the frequency f_A of A alleles in the diploid population as described in the introduction. To study how f_A changes over time, Fisher made certain assumptions about how the next generation of N individuals will be formed:

> [I]f a second generation of N individuals be now formed at random, the standard departure of f_A from its previous value will be $\sigma_{f_A} = \sqrt{f_A f_a / 2N}$. (Fisher, 1922a, p. 327)[9]

Fisher's statement reflected the assumptions of the binomial model: (i) the population size is finite and remains constant throughout its evolution; (ii) the second generation completely replaces the first; (iii) the second generation can be treated as a random sample drawn from the first; and (iv) the probability distribution of f_A is binomial. (i) and (ii) are indicated by the fact that Fisher considered the second generation of individuals whose number is equal to the population size. (iii) and (iv) are implicit in Fisher's claim that the second generation is "formed at random" and in his formula for the standard departure of f_A, which is related to the binomial distribution.

2.1. The binomial model and Fisher's statistical theory

Fisher's binomial model closely followed the concepts he developed in his contemporaneous work on mathematical statistics, such as population, sample, parameter, and statistic (Fisher, 1922b, 1925a, b). Table 1 provides an overview of our argument below.

In statistics, Fisher was concerned with estimation theory and introduced the concept of a parameter as a (typically unknown) quantity describing a population, the concept that forms a foundation of modern parametric statistical inference (Fisher, 1922b; Stigler, 1976; Hald, 2007). To estimate parameter values from observations, according to Fisher, we need to obtain a statistic from our observations. This procedure makes some assumptions about the population and treats observations as a random sample. As Fisher wrote:

> Any body of numerical observations, or qualitative data thrown into a numerical form as frequencies, may be interpreted as a random sample of some infinite hypothetical population of possible values. Problems of estimation arise when we know, or are willing to assume, the form of the frequency distribution of the population, as a mathematical function involving one or more unknown parameters, and wish to estimate the values of these parameters by means of the observational record available. A statistic may be defined as a function of the observations designed as an estimate of any such parameter. (Fisher, 1925b, p. 701; see also Fisher, 1925a, p. 3)

Suppose we toss a coin ten times to obtain a sample of coin tosses. In our sample, the observed frequency of heads may serve as a statistic estimating the (true but unknown) probability of heads. The value of this statistic depends on a particular random sample, so a statistic is a number that changes randomly in sampling experiments. The probability distribution of a statistic is called a sampling distribution, and the parameters of this distribution describe what Fisher referred to in the above passage as an "infinite hypothetical population of possible values [i.e., values of a statistic]." In the case of coin tosses, one such parameter is the probability of heads (and another is the number of tosses), and the sampling distribution of a statistic is binomial. Fisher discussed the binomial distribution along with other distributions in *Statistical methods for research workers* (Fisher, 1925a, pp. 65–66).

In the passage quoted above, Fisher referred to a hypothetical infinite population because he interpreted probabilities as frequencies. He said:

> The idea of an infinite hypothetical population is, I believe, implicit in all statements involving mathematical probability. If, in a Mendelian experiment, we say that the probability is one half that a mouse born of a certain mating shall be white, we must conceive of our mouse as one of an infinite population of mice which might have been produced by that mating. . . . Briefly, the hypothetical population is the conceptual resultant of the conditions which we are studying. The probability, like other statistical parameters, is a numerical characteristic of that population. (Fisher, 1925b, p. 700; see also Fisher, 1922b, p. 312)

If we are tossing a fair coin, the probability of obtaining heads is 0.5. In Fisher's statistical theory, this probability is a parameter describing the hypothetical infinite population of coin tosses.

Fisher's statistical concepts have counterparts in his binomial model. Recall that the model assumes that each generation of organisms is a

---

[6] These studies tend to focus on the relationship between Fisher's evolutionary thought and thermodynamics (Hodge, 1992; Depew & Weber, 1995, Ch. 10) or statistical mechanics (Morrison, 2000, pp. 214–224; Morrison, 2002; Plutynski, 2006; Winther, Giordano, Edge, & Nielsen, 2015). For exceptions to this trend, see Dietrich and Skipper (2013) and Winther et al. (2015, pp. 19–20).

[7] Feller (1949) also applied the theory of stochastic processes to evolution.

[8] Here are the journals with the year of founding in parentheses: *Journal of Theoretical Biology* (1961), *Journal of Applied Probability* (1964), *Advances in Applied Probability* (1969), and *Theoretical Population Biology* (1970).

[9] Fisher used different symbols: the lowercase n for population size, p and q for frequencies of A and a.

Table 1
Statistical concepts in Fisher's binomial model.

Statistical concepts	Statistical concepts in the binomial model
Population	Hypothetical infinite population of gametes; or theoretical population of allele frequencies
Sample	Organisms at each generation *t*; or $f_A(t)$
Population parameters	N and p_A ($= f_A(t)$, given neutrality)
Sample statistic	f_A
Sampling distribution	Binomial; characterized by population parameters
Standard error	$\sigma_{f_{A(t+1)}} = \sqrt{f_A(t) f_a(t)/2N}$ (for two alleles at a locus)

random sample drawn from the previous generation. In other words, just as Fisher's mouse in the above passage is a sample drawn from a hypothetical infinite population of potential mice ("mice which might have been produced by that mating"), organisms in one generation are a sample drawn from a hypothetical infinite population of potential organisms—or gametes.[10] Thus, the dynamics of Fisher's binomial model can be understood as follows (Fig. 1): In each generation, organisms produce a hypothetical infinite population of gametes. This population will have the same allele frequencies as the generation that produced it (assuming neutrality). From this population, a random sample of $2N$ gametes is drawn to produce N diploid organisms of the next generation. As a result of this sampling experiment, the allele frequency of the next generation may differ from that of the previous generation.

2.2. Evolution and the distribution of allele frequencies

Not only did Fisher use his statistical concepts to construct the binomial model, but he also used his theory of statistics as a mathematical foundation for using the binomial model to study evolution. In particular, his innovation was to treat the frequency f_A of A alleles in each generation as a statistic so that the sampling distribution of f_A can be interpreted as containing the information about how f_A may change over time.[11] In the binomial model, the sampling distribution of f_A is binomial because each of the $2N$ sampling experiments has only two outcomes (either A or a is selected), and the parameters of this distribution at generation t are the probability $p_A(t)$ of drawing an A allele and the sample size $2N$. These parameters characterize a hypothetical infinite population of possible allele frequencies. Now, given the assumptions of the binomial model, $p_A(t)$ is equal to $f_A(t)$, but because of the sampling experiments between generations, $f_A(t+1)$ may differ from $f_A(t)$. To analyze this possibility, Fisher said that the standard departure of f_A from its previous value is $\sigma_{f_A} = \sqrt{f_A f_a / 2N}$, by analogy with the standard error of a statistic.[12] In other words, $f_A(t+1)$ might differ from $f_A(t)$ as if due to sampling error.

Studying evolution through the change in the sampling distribution of f_A, Fisher focused on the stable form that this distribution should approach after a population evolved according to the binomial model for many generations. He found that the distribution would become flat (uniform), excluding the extremes corresponding to the extinction and fixation of an allele (Fisher, 1922a, pp. 326–330; Fisher, 1930, pp. 86–88). Once becoming flat, the distribution would become shorter and shorter in height as the random extinction of A gets more and more likely (i.e., the columns at the extremes become taller and taller).[13]

The rate at which the flattening of the distribution occurs is called the rate of decay of variability, because extinction of alleles means that a population loses its potential to change. This rate is relevant to assess Hagedoorn and Hagedoorn's (1921) claim about the relative importance of random extinction. Fisher (1922a, p. 330) argued that the rate of decay of variability is $1/4N$ and that it is too slow to be an important evolutionary influence. However, in correspondence, Wright pointed out an error in Fisher's argument, and they eventually agreed that the correct rate is $1/2N$ (Fisher, 1930, p. 87; Wright, 1931, pp. 119–121; Provine, 1986, pp. 243–260; Rosales, 2017).[14] As is well-known, even after this correction, Fisher and Wright disagreed about the importance of random drift in evolution.

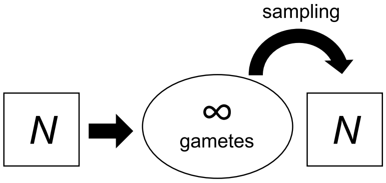

Fig. 1. Fisher's binomial model.

3. Malécot's reformulation of the binomial model

Fisher's construction and analysis of the binomial model were based on his theory of mathematical statistics. In the 1940s, however, the binomial model began to be formulated and studied with a new mathematical theory: the theory of Markov chains. We will show how this development had profound impacts on theoretical population genetics in terms of its directions of inquiry and a basic conception of evolution. To do that, we need to present the theory of Markov chains in some detail.

3.1. Markov chains

Let $X(t)$ be a variable that takes a certain value at time t. Consider the sequence of $X(t)$ for three time steps:

$$X(1),\quad X(2),\quad X(3).$$

Suppose this sequence has the following property. The probability that $X(3)$ takes a certain value depends on the value of $X(2)$, the similar dependency holds between $X(2)$ and $X(1)$, and given the value of $X(2)$, $X(3)$ is probabilistically independent of $X(1)$. In other words, the probability that $X(t+1)$ takes a certain value depends only on the value of $X(t)$. This property is called the Markov property, and A. A. Markov called the sequence of numbers with this property "numbers connected into a chain" (Markov, 1971 [1908], p. 552; see also Sheynin, 1989, pp. 364–370; Seneta, 1996).

In the theory of Markov chains, $X(t)$ is understood to be a random variable: a function defined on a sample space. The collection of all $X(t)$'s—that is, a collection of random variables indexed by parameter t's—is called a stochastic process (Khintchine, 1934, p. 604; Doob, 1934, p. 376), and the set of all possible values that each $X(t)$ can take is called the state space of a stochastic process. A stochastic process with the Markov property is a Markov process, and a Markov chain is a Markov process with a discrete state space.[15]

---

[10] Fisher (1925b, pp. 700–701) considered a hypothetical infinite population to be a certain limit of finite populations. For discussion of Fisher's idea, see Savage (1976, pp. 461–462), and for a philosophical study of probabilities and infinite populations in evolutionary theory, see Abrams (2006).

[11] For Fisher's own reflection on this move, see Fisher (1953, p. 515).

[12] For discussion of this concept in formulating the criteria of estimation, see Fisher (1922b, pp. 316–317; 1925b).

[13] That is, the distribution looks U-shaped (see Wright, 1931, p. 114).

[14] Wright (1931, pp. 110–111) also generalized Fisher's binomial model by introducing the concept of effective population size, and he also obtained the distributions of allele frequencies for the binomial model under a variety of conditions beyond complete neutrality, such as those including the influences of selection, mutation, and migration (Wright, 1931, pp. 134–137; see also Provine, 1986, pp. 239–242). In the 1930s and early 1940s, it was Wright, not Fisher, who did the most extensive study of the distribution of allele frequencies for the binomial model (see, e.g., Wright, 1937a, b; 1986 [1939]; 1945).

[15] A diffusion process is another type of Markov process, and it can have a continuous state space. For details, including other types of stochastic process,

Suppose we have a Markov chain with state space S. The transition probability of a Markov chain—the idea briefly explained at the beginning of this paper—can then be defined as follows:

$$p_{ij} = Pr\{X(t+1) = j | X(t) = i\}, \quad i, j \in S \tag{2}$$

where Pr(·) is a probability function. The exact form of Equation (2) of course depends on various assumptions about a Markov chain in question. Equation (1), introduced earlier, is just one example. Transition probabilities are probabilistic laws in the sense that they govern the behavior of a chain at every time step.

The theory of Markov chains studies various properties and behaviors of chains (Karlin & Taylor, 1975; Durrett, 2016). For example, Markov, (1971 [1908]) himself was interested in whether limit theorems of probability theory hold for chains.[16] Another important question concerns the asymptotic behavior of a chain: Given a certain initial condition, what, if any, can we say about the behavior of the chain after many time steps? In other words, what is the probability distribution of $X(t)$ as $t \to \infty$? We will shortly see that this question has a direct relevance to population genetics.

3.2. Malécot's reformulation of the binomial model

Until the 1930s, Markov's original work on chains was not widely known, and the topic of chains was somewhat neglected (Link, 2006; Mazliak, 2007, p. 162). Independently of Markov, Henri Poincaré, in 1912, analyzed chains that arise in the card shuffling problem he called “le battage des cartes” (Poincaré, 1912, p. 315). The problem was to study the probability that the order of cards in a deck changes from one to another after shuffling. In the 1920s, French and French-speaking mathematicians, such as Jacques Hadamard, Bohuslav Hostinský, and Maurice Fréchet, studied Poincaré’s shuffling problem and developed, in effect, the French version of the theory of chains (Bru, 2003; see also von Plato, 1994, p. 133; Link, 2006, p. 572).[17] By 1930, the French had learned about Markov's original work and the Russian work on stochastic processes, and in the 1930s in Paris, the Institut Henri Poincaré became an international center of research on the theory of Markov chains (Mazliak, 2007).

It was in Paris where Gustave Malécot studied probability theory, stochastic processes, Markov chains, and population genetics. From 1932 to 1935, he studied at the École Normale Supérieure, attending lectures by the leading French mathematicians Émile Borel, Élie Cartan, Maurice Fréchet, and Georges Darmois (Gillois, 2002). Subsequently, under Darmois, Malécot held a research fellowship at the Institut Henri Poincaré (Nagylaki, 1989, p. 254). Both Fréchet and Darmois taught Markov chains in the 1930s (Mazliak, 2007, p. 157, 166) and advised several outstanding doctoral students in probability theory, including Malécot (see Leloup, 2010).[18] Fréchet's systematic work on the theory of Markov chains during this time culminated in his 1938 treatise (Fréchet, 1938). It is important to note that advanced work on Markov chains was done in French in the 1930s because Malécot did not read German or Russian.[19] For Malécot's thesis, Darmois encouraged Malécot to study the work of Fisher and Wright in population genetics, and during his thesis research, Malécot discussed population genetics with Georges Teissier and Philippe L’Héritier (Malécot, 1939, p. 2).[20] In 1939, Malécot submitted his thesis entitled *Théorie mathématique de l'hérédité mendélienne généralisée* to the Faculty of Sciences at the University of Paris (Malécot, 1939). Using probability theory, Malécot's thesis turned Fisher’s (1918) theory of Mendelian heredity into a mathematically rigorous theory. Malécot thus became perhaps the first mathematician trained in probability theory, including the theories of stochastic processes and Markov chains, who specialized in theoretical population genetics.

In 1944, Malécot published a paper entitled, “Sur un problème de probabilités en chaîne que pose la génétique (On a problem of probabilities in chain posed by genetics)” (Malécot, 1944; for a commentary, see Seneta, 1974). In this brief paper, Malécot reformulated the binomial model as a Markov chain. Here is the key passage with the modern notion we adopted for Equation (1):

> Let us consider then the problem posed by Fisher [1930, p. 85f] and S. Wright [1931]: To study the probability $p_{ij}^{(t)}$ that in a finite and constant population of N individuals, a Mendelian gene A, appearing at a given moment in i loci on $2N$… appears t generations later in j loci on $2N$… Let us suppose first that the gene A does not confer, compared to its allelomorph a, any advantage or disadvantage on its carrier (no selection) and that it cannot transform into the other genes or appear from them (no mutation). Let us suppose finally that there is panmixia, that is to say that the genes A and a, which occupy the $2N$ loci in each generation, can be regarded as drawn at random from an infinitely large number of gametes A and a produced by the previous generation in the same proportions where A and a appear in its loci [i.e., the loci in the previous generation]. We then have:
>
> $$p_{ij}^{(1)} = \binom{2N}{j}\left(\frac{i}{2N}\right)^{j}\left(1-\frac{i}{2N}\right)^{2N-j} \quad [...]$$
>
> and the $p_{ij}^{(t)}$ are the probabilities in a simple and constant chain.” (Malécot, 1944, p. 380; our translation)

The passage clearly states the assumptions of the binomial model discussed in Section 2, and Malécot's equation is identical to the canonical formulation of the Wright-Fisher model, Equation (1). To use the terminology introduced in Section 3.1, $X(t)$ in Malécot's model represents the number of A alleles at a given generation, and the collection of all $X(t)$'s is a Markov chain whose probabilistic laws are given by Malécot's equation.

3.3. Stochastic processes and population genetics

Malécot's reformulation of the binomial model as a Markov chain opened a new direction of theoretical inquiries. For example, in the theory of Markov chains, it is common to arrange transition probabilities into a transition probability matrix **P**, which, for Malécot's model, looks like this:

$$\mathbf{P} = \begin{bmatrix} p_{00} & p_{01} & \cdots & p_{0,2N} \\ p_{10} & p_{11} & \cdots & p_{1,2N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{2N,0} & p_{2N,1} & \cdots & p_{2N,2N} \end{bmatrix}$$

(*footnote continued*)
see Karlin and Taylor (1975) and Durrett (2016).

[16] See Seneta (1996) for an interesting history behind Markov's interest in limit theorems and chains.

[17] As can be seen in Hadamard’s (1929) paper, the interest in the long-term properties of the card shuffling problem stemmed from its potential application to statistical mechanics. Hostinský was at a Czech university in Brno but wrote in French, including an early monograph on Markov chains (Hostinský, 1931). See Mazliak (2007) for details.

[18] The members of Malécot's thesis committee were Borel, Darmois, and Georges Valiron (see Malécot, 1939).

[19] In his interview, Malécot said that at the Institut Henri Poincaré, some of his classmates in the doctoral studies were influenced by the untranslated works of Kolmogorov and Khintchine on abstract topics in probability theory. But such topics did not interest Malécot, so Darmois suggested a more applied topic in population genetics. Malécot also said that he was “wrong not to read [Kolmogorov’s 1933 book], because the mathematical presentation is impeccable” (Bocquet-Appel, 1996, p. 106; our translation). An English translation of this book is Kolmogorov (1956).

[20] For the state of population genetics in France during the 1930s, see Gayon and Veuille (2001).

P contains essential information about the dynamics of Malécot's model, and importantly, it also allowed theoreticians to rigorously study the complete dynamics of Fisher's original model. If we know the initial state of the system, we can know the probability that the system will be in state j after t time steps by raising the matrix to the tth power, $\mathbf{P}^t$. Note that **P** itself does not change with time.[21] So after the introduction of the Markov-chain formulation, the theoretical study of drift has concentrated on the properties of this matrix.

One result of this new inquiry concerned the rate of decay of variability. In the theory of Markov chains, a state from which the system cannot leave once entered is called an absorbing state. In Malécot's model, without mutation, fixation and loss of an allele are absorbing states, and when the population enters these states, its variability is lost. The rate at which the system will reach absorbing states, then, is the rate of decay of variability. For Malécot's model, Feller (1951) showed that the probability of fixation is $X(0)/2N$, that is, the initial frequency of A alleles. He also argued that for large t (i.e., the number of generations passed), the rate of decay of variability depends on what mathematicians call the eigenvalues, latent roots, or characteristic values of **P** and that the rate is approximately $1/2N$ (for explanations of the mathematical derivations, see Ewens, 1969, pp. 29–32; 2004, pp. 21–23). Feller's finding agrees with the conclusion that Fisher and Wright reached by different methods. Malécot (1944) also arrived at the same conclusion by yet another method.

Another result concerned the stable form of the distribution of allele frequencies. Given Malécot's model, what needs to be found is the probability distribution of $X(t)$ as $t \to \infty$. Not only did Malécot (1944) show that the asymptotic probability distribution is uniform, just as Fisher (1922a, 1930) and Wright (1931) had argued, but he also showed, more rigorously than Fisher and Wright, that any distribution of $X(t)$, with $X(0) \neq 0, 2N$, will converge to the uniform distribution.

Beyond these results, the Markov-chain model turned out to be difficult to analyze. However, Malécot (1944, 1945, 1948) and Feller (1951) argued that with appropriate assumptions, the dynamics of the Markov-chain model can be approximated by a diffusion process with continuous state space. With this approximation, the mathematical analysis can be done with the powerful methods of differential and integral equations. Now, to study the distribution of allele frequencies, Fisher had used a partial differential equation, like the one used to study heat diffusion (Fisher, 1922a, p. 327), and Wright had used integral equations (e.g., Wright, 1931, p. 113). What Malécot and Feller pioneered was the use of a diffusion process as an approximation to a discrete Markov chain.[22]

Thus, Malécot's reformulation introduced a new mathematical theory to explore the dynamics of Fisher's binomial model. But, in our view, Malécot's reformulation also had a more general significance: it introduced a new conception of evolution that did not depend on the Fisherian statistical concepts. To understand this conception, it is important to note that the theory of Markov chains and the general theory of stochastic processes were both conceived as a mathematical theory about a physical system whose state changes randomly (see, e.g., Kolmogorov, 1931, p. 415; Fréchet, 1938, p. 23; Chung, 1960, p. 4). As Feller explained:

> Conceptually, a Markov process is the probabilistic analogue of the processes of classical mechanics, where the future development is completely determined by the present state and is independent of the way in which the present state has developed … In stochastic processes the future is never uniquely determined, but we have at least probability relations enabling us to make predictions. For the Markov chains … it is clear that probability relations relating to the future depend on the present state, but not on the manner in which the present state has emerged from the past. (Feller, 1950, p. 337; see also von Plato, 1994, p. 209)

Referring to the way in which the future of a Markov process is probabilistically determined, Kolmogorov wrote:

> A physical process (a change of a certain physical system) is called *stochastically determined* if, knowing a state X_0 of the system at a certain moment of time t_0 we also know the probability distribution for all the states X of this system at the moments $t > t_0$. (Kolmogorov, 1931, p. 415; the English translation is from Shiryayev, 1992, p. 62)

Taking Kolmogorov's and Feller's views together, we can see that in Malécot's model, the evolutionary change in the number of alleles in a population is conceived as a stochastically determined process whose time evolution is well defined, in analogy to the motion of point particles in classical mechanics. Let us call this conception of evolution *the stochastic dynamical conception*.[23] This is different from the Fisherian statistical conception, which analogizes evolution to a series of sampling experiments.

4. The Markovian path of theoretical population genetics

After Malécot’s (1944) work, a field of theoretical research based on the stochastic dynamical conception emerged, leading to what we call the Markovian path of theoretical population genetics.[24]

In the early 1950s, Malécot's Markov-chain model was introduced to the Anglophone audience by Feller, in his famous textbook on probability (Feller, 1950, p. 315) and in his influential paper that showed the utility of diffusion theory in the analysis of the Markov-chain model (Feller, 1951, pp. 231–232; Nagylaki, 1989, p. 258). The rest of the 1950s and the 1960s saw a rapid development of the Markovian path. For example, Malécot's Markov-chain model was discussed in introductory textbooks on probability theory and stochastic processes (Feller, 1950, p. 315; Bartlett, 1955, pp. 120–123), and the theory of stochastic processes, including Markov chains, was in turn discussed in introductory textbooks on population genetics (Li, 1955, Ch. 22; Kempthorne, 1957, Ch. 4). In a series of papers starting in the mid-1950s, Kimura used diffusion theory to solve a number of theoretical problems that originally arose from the Markov-chain model (Kimura, 1954, 1955a, b; reviewed in Kimura, 1964; see also Watterson, 1996). In 1958, P. A. P. Moran introduced a new Markov-chain model that described a haploid population with overlapping generations (Moran, 1958).

By the 1960s, then, the Markovian path of theoretical population genetics was established as a field in its own right. Time was ripe for Moran to introduce Markov chains in his textbook on theoretical population genetics (Moran, 1962) and for his students, Geoffrey Watterson and Warren Ewens, to write their dissertations on the application

---

[21] This property is called stationarity or time-homogeneity. The claim about $\mathbf{P}^n$ follows from what is now known as the Chapman-Kolmogorov equation, which is fundamental to the general theory of stochastic processes that Kolmogorov (1931) initiated (see Chung, 1960, pp. 1–9; Durrett, 2016, p. 11).

[22] Before Malécot, mathematicians had studied diffusion processes in relation to Markov chains (e.g., Fréchet, 1938; Khintchine, 1933; Kolmogorov, 1931, 1933). Historically, Kolmogorov (1935) was the first to use diffusion theory to study a problem in population genetics. In 1944, Malécot was not aware of Kolmogorov's paper; in fact, it was not widely known until Kolmogorov sent his paper to Wright, who introduced it to the Anglophone audience (Wright, 1945). Kolmogorov's paper is reprinted in Shiryayev (1992, pp. 179–181).

[23] Sarkar develops an interpretation of evolutionary theory that he calls “the stochastic dynamical interpretation” (Sarkar, 2011, p. 464). What we called the stochastic dynamical conception of evolution is similar to what Sarkar calls “the minimal dynamical interpretation” (Sarkar, 2011, p. 464), according to which evolving populations are dynamical systems. Sarkar traces this interpretation back to Lewontin (1974) and Hirsch and Smale (1974) (see Sarkar, 2011, p. 449, p. 461).

[24] The Markovian path, though, does not have the Markov property.

of the theory of stochastic processes to population genetics (Ewens, 1963; Watterson, 1960). A new textbook on stochastic processes presented the Markov-chain formulation of the binomial model as one of the basic examples of Markov chains (Karlin, 1966, pp. 37–39), although it did not cite Malécot's work. Kimura's (1964) influential review of the diffusion theory in population genetics appeared in the second issue of newly founded *Journal of Applied Probability*. Ewens's textbook and the English translation of Malécot's 1948 book both appeared in 1969 (Ewens, 1969; Malécot, 1969). New Markov-chain models in population genetics were also introduced in this decade (Chia & Watterson, 1969; Karlin & McGregor, 1965).

In all these works, the binomial model was understood to be a Markov chain, and in the English-language publications, both the Markov chain and its diffusion approximation were typically attributed to Fisher and Wright.[25] Towards the end of the 1960s, the now standard name for Malécot's Markov-chain model, "the Wright-Fisher model," began to appear. For example, in one paper, Karlin used both "the Fisher-Wright random sampling model" and "a model based on the classical Wright-Fisher random sampling formulation" (Karlin, 1968, p. 446, p. 454), and Karlin and McGregor used "the Wright-Fisher binomial sampling model" (Karlin & McGregor, 1968, p. 144).

Population geneticists became increasingly accustomed to the idea that the Wright-Fisher model was originally formulated by Wright and Fisher—in this order—as a Markov chain.[26] Thus, in his review of an edited collection of reprints entitled *Stochastic models in population genetics* (Li, 1977), A. W. F. Edwards pointed out that the collection reinforced the customary thought, and he then went on to correct it: Fisher introduced the binomial model, and Wright used the same model but introduced the effective population size and the integral equation methods for studying the distribution of allele frequencies (Edwards, 1978, p. 240; see also footnote 14 above). Edwards also noted that Malécot (1944) was the first to apply the theory of Markov chains to Fisher's model and lamented the omission of Malécot's paper in the collection (Edwards, 1978, p. 241).[27] From our point of view, Edwards's critical review reflects how canonical the Wright-Fisher model as a Markov chain has become in theoretical population genetics.

5. Philosophical implications

As we have argued, the historical origins of the stochastic theory of population genetics include Fisher's binomial model, Malécot's reformulation of the model as a Markov chain, and the emergence of the Markovian path as a field of theoretical population genetics based on the stochastic dynamical conception of evolution. Understanding these origins and their historical and conceptual connections sheds new light on philosophical research on drift by generating unexplored questions, some of which we formulate below.[28]

Philosophical research on drift tends to rely heavily on the concept of random sampling. For example, some philosophers argue that drift is a physical sampling process (e.g., Beatty, 1984; Hodge, 1987; Millstein, 2002), while others argue that drift is sampling error, that is, a statistical effect (e.g., Walsh, Lewens, & Ariew, 2002). Philosophers also use the concept of random sampling to explain the importance of (small) population size for drift: small population size is analogous to small sample size, which increases the amount of sampling error (e.g., Gildenhuys, 2009; Reisman & Forber, 2005; Walsh et al., 2002).

Of course, biologists today use random sampling and sampling error to illustrate drift, so the use of the sampling concepts in philosophical discussion is legitimate. Historically, these sampling concepts entered theoretical population genetics through Fisher's binomial model whose mathematical foundation was the theory of mathematical statistics. As we have shown, with the emergence of the Markovian path, the theory of stochastic processes became a foundation for theoretical population genetics. Given this new foundation, would the sampling concepts remain essential for understanding drift?

The historical development of the Markovian path suggests that the sampling concepts may not be so essential. It is true that Malécot (1944) assumed random sampling of gametes in his presentation of the Markov-chain model (see the relevant passage quoted in Section 3.2). But in the 1960s theoreticians started to generalize Malécot's model by relaxing its assumptions, including the one about random sampling.[29] A crux of this development came shortly after the historical period we discussed above: In 1974, Chris Cannings showed that the Wright-Fisher model, Moran's model, and other drift models introduced in the 1960s are all instances of a general Markov-chain model of drift, and crucially, this general model does not include any assumption about random sampling (Cannings, 1974, pp. 260–269).[30] In light of the development of the Markovian path, then, how are we to understand drift?

The Markovian path raises another philosophical question. As we noted in Section 3.2, the Markovian path made explicit the theoretical relationship between the Markov-chain model and a diffusion process with continuous state space: the latter is an approximation to the former. In practice, this approximation involves delicate rescaling of time and the state space of the Markov-chain model (for details, see Feller, 1951, pp. 239–240; Ewens, 2004, Chs. 4–5). Roughly, a diffusion process is assumed to describe *continuous* change in allele frequency (i.e., $X(t)/2N$) over a certain time interval, and to this continuous-time continuous-state process, in some respects, the behavior of the Markov-chain model is thought to converge in the limit as $N \to \infty$. Thus, as the Markovian path developed, theoreticians have realized that drift is important even in large populations, but the received view among philosophers is sill that drift becomes unimportant as population size increases. While this view makes sense when one is thinking about drift by analogy with sampling experiments like coin tosses, at least it deserves to be reconsidered in light of the Markovian path. How are we to understand drift if it can be important in large populations?[31]

Finally, the stochastic dynamical conception may allow us to reformulate a widely discussed question about the theoretical character of drift. Following Elliot Sober's (1984) influential characterization of evolutionary theory as a theory of forces, philosophers have debated whether drift can be seen as a force analogous to a Newtonian force (e.g., Matthen & Ariew, 2002; Stephens, 2004; Walsh et al., 2002). The

---

[25] Before the Markov-chain model was attributed to both Fisher and Wright, it was sometimes referred to as "the Wright model" or "Wright's model" (e.g., Chia & Watterson, 1969, p. 232; Ewens, 1969, p. 28; Karlin & McGregor, 1964, p. 246). We do not know who first used the now standard name of the model or why Wright's name came before Fisher's or why Malécot's name was never attached to the model. But the "Wright-Fisher" nomenclature might reflect two facts. First, Wright was undoubtedly *the* major theorist and expositor of the binomial model before textbooks on population genetics appeared in the 1950s and 1960s. Second, in the late 1940s, Fisher appeared to disown the binomial model by dubbing genetic drift "the Sewall Wright effect" and strongly criticizing Wright for considering drift to be significant in explaining the evolution of natural populations (Fisher & Ford, 1947, 1950; see also Wright, 1948).

[26] Although this idea is historically inaccurate, it is not conceptually incoherent since Malécot's Markov-chain model can be seen as a mathematically more rigorous version of Fisher's binomial model.

[27] Interestingly, however, Edwards also referred to Fisher's original model as "the *chain*-binomial model" (Edwards, 1978, p. 240; emphasis added).

[28] For reviews on the philosophical debates over drift, see Plutynski (2007), Otsuka (2016), and Millstein (2017). The philosophical questions about drift are part of the broader literature on chance in evolutionary biology (see, e.g., Ramsey & Pence, 2016).

[29] See the works cited in Section 4.

[30] The historical significance of the Cannings model and the development of the stochastic theory of population genetics in the 1970s and 1980s will be the topic of a sequel to the present paper.

[31] Millstein (2017) argues that her account of drift allows for its occurrence in an infinite population.

Fisherian statistical conception dominates this debate, but the stochastic dynamical conception offers a more natural home to the force metaphor. Following this conception, diffusion theory has been used to study the stochastic dynamics of evolving populations under the influence of interacting evolutionary factors, and in this context, theoreticians think about vectors moving in opposite or the same directions (see Kimura's work cited in Section 4). Is the force metaphor appropriate in this theoretical context? Why or why not?[32]

6. Conclusion

The Wright-Fisher model, understood as a Markov chain, is the basic model of the stochastic theory of population genetics. This model is a product of the historical development of theoretical population genetics from Fisher's 1922 paper to the widespread use of the theory of Markov chains in the 1960s. Fisher's binomial model was founded on his theory of mathematical statistics; his basic statistical concepts—population, sample, parameter, statistic, sampling distribution—all appeared in the binomial model. In using this model, Fisher saw stochastic changes in a population of organisms as changes in the population of allele frequencies, that is, the sampling distribution of allele frequencies.

In 1944, Fisher's binomial model was reformulated as a Markov chain by Malécot, who had studied probability theory, stochastic processes, and Markov chains when Paris was a major center of research in these areas. Malécot conceived changes in a population of organisms as changes in the state of a dynamical system governed by probabilistic laws. As Malécot's foundational reformulation became better known in the 1950s and 1960s, the Markovian path of theoretical population genetics emerged. This path developed primarily in the community of mathematical population geneticists, including Feller, Kimura, Moran, Karlin, Watterson, and Ewens. When we see the Wright-Fisher model as a basic model of random genetic drift, we are witnessing the result of the migration of people and ideas from mathematics to theoretical population genetics.

Declaration of competing interest

None.

Acknowledgements

We especially thank A. W. F. Edwards and Warren Ewens for a longstanding correspondence on the topics of this paper. We thank Sarah Otto and Nick Barton for discussions and comments. We acknowledge helpful comments by two anonymous referees.

References

Abrams, M. (2006). Infinite populations and counterfactual frequencies in evolutionary theory. *Studies in History and Philosophy of Biological and Biomedical Sciences, 37*, 256–268.
Adams, M. B. (1968). The founding of population genetics: Contributions of the Chetverikov school 1924–1934. *Journal of the History of Biology, 1*, 23–39.
Adams, M. B. (1970). Towards a synthesis: Population concepts in Russian evolutionary thought, 1925–1935. *Journal of the History of Biology, 3*, 107–129.
Bartlett, M. S. (1955). *An introduction to stochastic processes with special reference to methods and applications.* Cambridge: Cambridge University Press.
Beatty, J. (1984). Chance and natural selection. *Philosophy of Science, 51*, 183–211.
Beatty, J. (1987). Dobzhansky and drift: Facts, values, and chance in evolutionary biology. In L. Krüger, G. Gigerenzer, & M. S. Morgan (Eds.). *The probabilistic revolution: Volume 2, Ideas in the sciences* (pp. 271–311). Cambridge, MA: MIT Press.
Beatty, J. (1992). Random drift. In E. F. Keller, & E. A. Lloyd (Eds.). *Keywords in evolutionary biology* (pp. 237–281). Cambridge, MA: Harvard University Press.
Bocquet-Appel, J.-P. (1996). Interview de Gustave Malécot (mai 1993, avril 1994). *Bulletin et Mémoires de la Société d'Anthropologie de Paris, 8*, 105–114.
Bru, B. (2003). Souvenirs de Bologne. *Journal de la Société Française de Statistique, 144*, 135–226.
Burian, R. M., & Gayon, J. (1999). The French school of genetics: From physiological and population genetics to regulatory molecular genetics. *Annual Review of Genetics, 33*, 313–349.
Cannings, C. (1974). The latent roots of certain Markov chains arising in genetics: A new approach, I. Haploid models. *Advances in Applied Probability, 6*, 260–290.
Charlesworth, B., & Charlesworth, D. (2017). Population genetics from 1966–2016. *Heredity, 118*, 2–9.
Chia, A. B., & Watterson, G. A. (1969). Ecological effects on the rate of genetic evolution. I. Constant size populations with two genotypes. *Journal of Applied Probability, 6*, 231–248.
Chung, K. L. (1960). *Markov chains with stationary transition probabilities.* Berlin: Springer-Verlag.
Crow, J. F. (1987). Population genetics history: A personal view. *Annual Review of Genetics, 21*, 1–22.
Depew, D. J., & Weber, B. H. (1995). *Darwinism evolving: Systems dynamics and the genealogy of natural selection.* Cambridge, MA: MIT Press.
Dietrich, M. R., & Skipper, R. A., Jr. (2013). R. A. Fisher and the foundations of statistical biology. In O. Harman, & M. R. Dietrich (Eds.). *Outsider scientists: Routes to innovation in biology* (pp. 147–160). Chicago: University of Chicago Press.
Doob, J. L. (1934). Stochastic processes and statistics. *Proceedings of the National Academy of Sciences of the United States of America, 20*, 376–379.
Dung, S. K., Lopéz, A., Baragan, E. L., Reyes, R.-J., Thu, R., Castellanos, E., et al. (2019). Illuminating women's hidden contribution to historical theoretical population genetics. *Genetics, 211*, 363–366.
Durrett, R. (2016). *Essentials of stochastic processes* (3rd ed.). New York: Springer.
Dynkin, E. B. (1989). Kolmogorov and the theory of Markov processes. *Annals of Probability, 17*, 822–832.
Edwards, A. W. F. (1978). Review of *Stochastic models in population genetics*. *Heredity, 41*, 239–242.
Edwards, A. W. F. (2001). Darwin and Mendel united: The contributions of Fisher, Haldane, and Wright, up to 1932. In E. C. R. Reeve (Ed.). *Encyclopedia of genetics* (pp. 77–83). London: Fitzroy Dearborn.
Epperson, B. K. (1999). Gustave Malécot, 1911–1998: Population genetics founding father. *Genetics, 152*, 477–484.
Ewens, W. J. (1963). *Stochastic processes in population. genetics* Canberra, Australia: Australian National University (Unpublished doctoral dissertation).
Ewens, W. J. (1969). *Population genetics.* London: Methuen.
Ewens, W. J. (2004). *Mathematical population genetics: I. Theoretical introduction* (2nd ed.). New York: Springer.
Feller, W. (1949). On the theory of stochastic processes, with particular reference to applications. In J. Neyman (Ed.). *Proceedings of the Berkeley symposium on mathematical statistics and probability* (pp. 403–432). Berkeley: University of California Press.
Feller, W. (1950). *An introduction to probability theory and its applications*, *Vol. 1*. New York: Wiley.
Feller, W. (1951). Diffusion processes in genetics. In J. Neyman (Ed.). *Proceedings of the second Berkeley symposium on mathematical statistics and probability* (pp. 227–246). Berkeley: University of California Press.
Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh, 52*, 399–433.
Fisher, R. A. (1922a). On the dominance ratio. *Proceedings of the Royal Society of Edinburgh, 42*, 321–341.
Fisher, R. A. (1922b). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London A, 222*, 309–368.
Fisher, R. A. (1925a). *Statistical methods for research workers.* Edinburgh: Oliver and Boyd.
Fisher, R. A. (1925b). Theory of statistical estimation. *Proceedings of the Cambridge Philosophical Society, 22*, 700–725.
Fisher, R. A. (1930). *The genetical theory of natural selection.* Oxford: Clarendon Press.
Fisher, R. A. (1953). Croonian lecture: Population genetics. *Proceedings of the Royal Society of London Series B Biological Sciences, 141*, 510–523.
Fisher, R. A., & Ford, E. B. (1947). The spread of a gene in natural conditions in a colony of the moth *Panaxia dominula L. Heredity, 1*, 143–174.
Fisher, R. A., & Ford, E. B. (1950). The "Sewall Wright effect". *Heredity, 4*, 117–119.
Fréchet, M. (1938). Recherches théoriques modernes sur le calcul des probabilités. *Méthode des fonctions arbitraires. Théorie des événements en chaîne dans le cas d'un nombre fini d'états possible: Vol. 2*. Paris: Gauthier.
Gayon, J., & Veuille, M. (2001). The genetics of experimental populations: L'Heritier and teissier's population cages. In R. Singh, K. Krimbas, J. Beatty, & D. Paul (Eds.). *Thinking about evolution: Historical, philosophical, and political perspectives* (pp. 77–102). Cambridge: Cambridge University Press.
Gildenhuys, P. (2009). An explication of the causal dimension of drift. *The British Journal for the Philosophy of Science, 60*, 521–555.
Gillois, M. (2002). The scientific work of Gustave Malécot (1911–1998): Our common heritage. In M. Slatkin, & M. Veuille (Eds.). *Modern developments in theoretical population genetics: The legacy of Gustave Malécot* (pp. 7–19). Oxford: Oxford University Press.
Hadamard, J. (1929). Sur le battage des cartes et ses relations avec la mécanique statistique. *Atti del Congresso Internazionale dei Matematici: Bologna, 3–10 Settembre 1928, 5*, 133–139.
Hagedoorn, A. L., & Hagedoorn, A. C. (1921). *The relative value of the processes causing evolution.* The Hague: Martinus Nijhoff.
Hald, A. (2007). *A history of parametric statistical inference from Bernoulli to Fisher, 1713–1935.* New York: Springer.
Hartl, D. L., & Clark, A. G. (2007). *Principles of population genetics* (4th ed.). Sunderland, MA: Sinauer.
Hirsch, M. W., & Smale, S. (1974). *Differential equations, dynamical systems, and linear*

---

[32] Pence (2017) evaluates the force metaphor as applied to drift by drawing on the use of the same metaphor in studies of Brownian motion, which is a diffusion process in the theory of stochastic processes.

algebra. New York: Academic Press.

Hodge, M. J. S. (1987). Natural selection as a causal, empirical, and probabilistic theory. In L. Kruger, G. Gigerenzer, & M. S. Morgan (Eds.). *The probabilistic revolution: Volume 2, Ideas in the sciences* (pp. 233–270). Cambridge, MA: MIT Press.

Hodge, M. J. S. (1992). Biology and philosophy (including ideology): A study of Fisher and Wright. In S. Sarkar (Ed.). *The founders of evolutionary genetics: A centenary reappraisal* (pp. 231–293). Dordrecht: Kluwer.

Hodge, J. (2011). Darwinism after Mendelism: The case of Sewall Wright's intellectual synthesis in his shifting balance theory of evolution (1931). *Studies in History and Philosophy of Biological and Biomedical Sciences, 42*, 30–39.

Hostinský, B. (1931). *Méthodes générales du calcul des probabilités.* Paris: Gauthier-Villars.

Ishida, Y. (2009). Sewall Wright and Gustave Malécot on isolation by distance. *Philosophy of Science, 76*, 784–796.

Ishida, Y. (2017). Sewall Wright, shifting balance theory, and the hardening of the modern synthesis. *Studies in History and Philosophy of Biological and Biomedical Sciences, 61*, 1–10.

Karlin, S. (1966). *A first course in stochastic processes.* New York: Academic Press.

Karlin, S. (1968). Rates of approach to homozygosity for finite stochastic models with variable population size. *The American Naturalist, 102*, 443–455.

Karlin, S., & McGregor, J. (1964). On some stochastic models in genetics. In J. Gurland (Ed.). *Stochastic models in medicine and biology* (pp. 245–279). Madison: University of Wisconsin Press.

Karlin, S., & McGregor, J. (1965). Direct product branching processes and related induced markoff chains: I. Calculations of rates of approach to homozygosity. In J. Neyman, & L. M. LeCam (Eds.). *Bernoulli, Bayes, Laplace: Anniversary volume* (pp. 111–145). New York: Springer.

Karlin, S., & McGregor, J. (1968). Rates of probabilities of fixation for two locus random mating finite populations without selection. *Genetics, 58*, 141–159.

Karlin, S., & Taylor, H. M. (1975). *A first course in stochastic processes* (2nd ed.). New York: Academic Press.

Kempthorne, O. (1957). *An introduction to genetic statistics.* New York: Wiley.

Khintchine, A. (1933). *Asymptotische gesetze der Wahrscheinlichkeitsrechnung.* Berlin: Julius Springer.

Khintchine, A. (1934). Korrelationstheorie der stationären stochastischen Prozesse. *Mathematische Annalen, 109*, 604–615.

Kimura, M. (1954). Process leading to quasi-fixation of genes in natural populations due to random fluctuation of selection intensities. *Genetics, 39*, 280–295.

Kimura, M. (1955a). Solution of a process of random genetic drift with a continuous model. *Proceedings of the National Academy of Sciences of the United States of America, 41*, 144–150.

Kimura, M. (1955b). Stochastic processes and distribution of gene frequencies under natural selection. *Cold Spring Harbor Symposia on Quantitative Biology, 10*, 33–53.

Kimura, M. (1964). Diffusion models in population genetics. *Journal of Applied Probability, 1*, 177–232.

Kolmogorov, A. N. (1931). Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Mathematische Annalen, 104*, 415–458.

Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Berlin: Springer.

Kolmogorov, A. N. (1935). Deviations from Hardy's formulas under partial isolation. *Comptes Rendus de l'Académie des Sciences de l'URSS, 3*, 129–132.

Kolmogorov, A. N. (1956). *Foundations of the theory of probability* (2nd English ed., N. Morrison, Trans.). New York: Chelsea.

Leloup, J. (2010). Dissertations on probability in Paris in the 1930s. *Journal Électronique d'Histoire des Probabilités et de la Statistique, 6*, 1–26.

Lewontin, R. C. (1974). *The genetic basis of evolutionary change.* New York: Columbia University Press.

Li, C. C. (1955). *Population genetics.* Chicago: University of Chicago Press.

Li, W.-H. (Ed.). (1977). *Stochastic models in population genetics*. Stroudsburg, PA: Dowden, Hutchinson and Ross.

Link, D. (2006). Chains to the West: Markov's theory of connected events and its transmission to Western Europe. *Science in Context, 19*, 561–589.

Malécot, G. (1939). *Théorie mathématique de l'hérédité mendélienne généralisée.* Paris: J. Guilhot.

Malécot, G. (1944). Sur un problème de probabilités en chaîne que pose la génétique. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences Paris, 219*, 379–381.

Malécot, G. (1945). La diffusion des gènes dans une population mendélienne. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences Paris, 221*, 340–342.

Malécot, G. (1948). *Les mathématiques de l'hérédité.* Paris: Masson et Cie.

Malécot, G. (1969). *The mathematics of heredity (D. M. Yermanos, Trans).* San Francisco: W. H. Freeman.

Markov, A. A. (1971 [1908]). Extension of the limit theorems of probability theory to a sum of variables connected in a chain. *Dynamic probabilistic systems, volume I: Markov models* (pp. 552–576). New York: Wiley.

Matthen, M., & Ariew, A. (2002). Two ways of thinking about fitness and natural selection. *Journal of Philosophy, 99*, 55–83.

Mazliak, L. (2007). On the exchanges between Wolfgang Doeblin and Bohuslav Hostinský. *Revue d'Histoire des Mathématiques, 13*, 155–180.

Millstein, R. L. (2002). Are random drift and natural selection conceptually distinct? *Biology and Philosophy, 17*, 33–53.

Millstein, R. L. (2017). Genetic drift. In E. N. Zalta (Ed.). *The Stanford encyclopedia of philosophy* (Fall 2017 Edition). https://plato.stanford.edu/archives/fall2017/entries/genetic-drift/, Accessed date: 12 May 2019.

Moran, P. A. P. (1958). Random processes in genetics. *Proceedings of the Cambridge Philosophical Society, 54*, 60–71.

Moran, P. A. P. (1962). *The statistical processes of evolutionary theory.* Oxford: Clarendon Press.

Morrison, M. (2000). *Unifying scientific theories: Physical concepts and mathematical structures.* Cambridge: Cambridge University Press.

Morrison, M. (2002). Modelling populations: Pearson and Fisher on mendelism and biometry. *The British Journal for the Philosophy of Science, 53*, 39–68.

Nagylaki, T. (1989). Gustave Malécot and the transition from classical to modern population genetics. *Genetics, 122*, 253–268.

Neuhauser, C. (2001). Mathematical models in population genetics. In D. J. Balding, M. Bishop, & C. Cannings (Eds.). *Handbook of statistical genetics* (pp. 153–177). New York: Wiley.

Otsuka, J. (2016). A critical review of the statisticalist debate. *Biology and Philosophy, 31*, 459–482.

Pence, C. (2017). Is genetic drift a force? *Synthese, 194*, 1967–1988.

von Plato, J. (1994). *Creating modern probability: Its mathematics, physics and philosophy in historical perspective.* Cambridge: Cambridge University Press.

Plutynski, A. (2005). Parsimony and the Fisher-Wright debate. *Biology and Philosophy, 20*, 697–713.

Plutynski, A. (2006). What was Fisher's fundamental theorem of natural selection and what was it for? *Studies in History and Philosophy of Biological and Biomedical Sciences, 37*, 59–82.

Plutynski, A. (2007). Drift: A historical and conceptual overview. *Biological Theory, 2*, 156–167.

Poincaré, H. (1912). *Calcul des probabilités* (2nd ed.). Paris: Gauthier-Villars.

Provine, W. B. (1971). *The origins of theoretical population genetics.* Chicago: University of Chicago Press.

Provine, W. B. (1978). The role of mathematical population geneticists in the evolutionary synthesis of the 1930s and 1940s. *Studies in History of Biology, 2*, 167–192.

Provine, W. B. (1985). The R. A. Fisher–Sewall Wright controversy. In R. Dawkins, & M. Ridley (Eds.). *Oxford surveys in evolutionary biology* (pp. 197–219). Oxford: Oxford University Press.

Provine, W. B. (1986). *Sewall Wright and evolutionary biology.* Chicago: University of Chicago Press.

Provine, W. B. (1990). Discussion: Population genetics. *Bulletin of Mathematical Biology, 52*, 201–207.

Ramsey, G., & Pence, C. H. (Eds.). (2016). *Chance in evolution.* Chicago: University of Chicago Press.

Reisman, K., & Forber, P. (2005). Manipulation and the causes of evolution. *Philosophy of Science, 72*, 1113–1123.

Rosales, A. (2017). Theories that narrate the world: Ronald A. Fisher's mass selection and Sewall Wright's shifting balance. *Studies in History and Philosophy of Science, 62*, 22–30.

Sarkar, S. (2011). Drift and the causes of evolution. In P. M. Illari, F. Russo, & J. Williamson (Eds.). *Causality in the sciences* (pp. 445–469). Oxford: Oxford University Press.

Savage, L. J. (1976). On rereading R. A. Fisher. *Annals of Statistics, 4*, 441–500.

Seneta, E. (1974). A note on the balance between random sampling and population size (on the 30th anniversary of G. Malécot's paper). *Genetics, 77*, 607–610.

Seneta, E. (1996). Markov and the birth of chain dependence theory. *International Statistical Review, 64*, 255–263.

Sheynin, O. B. (1989). A. A. Markov's work on probability. *Archive for History of Exact Sciences, 39*, 337–377.

Shiryayev, A. N. (Ed.). (1992). *Selected works of A. N. Komogorov: Volume II, probability theory and mathematical statistics (G. Lindquist, Trans.)*. Dordrecht: Springer.

Skipper, R. A. (2002). The persistence of the R. A. Fisher-Sewall Wright controversy. *Biology and Philosophy, 17*, 341–367.

Slatkin, M., & Veuille, M. (Eds.). (2002). *Modern developments in theoretical population genetics: The legacy of Gustave Malécot.* Oxford: Oxford University Press.

Sober, E. (1984). *The nature of selection: Evolutionary theory in philosophical focus.* Cambridge, MA: MIT Press.

Stephens, C. (2004). Selection, drift, and the "forces" of evolution. *Philosophy of Science, 71*, 550–570.

Stigler, S. M. (1976). Discussion of Savage (1976). *Annals of Statistics, 4*, 498–500.

Turner, J. R. G. (1987). Random genetic drift, R. A. Fisher, and the Oxford school of ecological genetics. In L. Krüger, G. Gigerenzer, & M. S. Morgan (Eds.). *The probabilistic revolution: Volume 2, Ideas in the sciences* (pp. 313–354). Cambridge, MA: MIT Press.

Walsh, D. M., Lewens, T., & Ariew, A. (2002). The trials of life: Natural selection and random drift. *Philosophy of Science, 69*, 452–473.

Watterson, G. A. (1960). *Probability theory applied to genetic. populations* Canberra, Australia: Australian National University (Unpublished doctoral dissertation).

Watterson, G. A. (1996). Motoo Kimura's use of diffusion theory in population genetics. *Theoretical Population Biology, 49*, 154–188.

Winther, R. G., Giordano, R., Edge, M. D., & Nielsen, R. (2015). The mind, the lab, and the field: Three kinds of populations in scientific practice. *Studies in History and Philosophy of Biological and Biomedical Sciences, 52*, 12–21.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics, 16*, 97–159.

Wright, S. (1937a). The distribution of gene frequencies in populations. *Science, 85*, 504.

Wright, S. (1937b). The distribution of gene frequencies in populations. *Proceedings of the National Academy of Sciences of the United States of America, 23*, 307–320.

Wright, S. (1945). The differential equation of the distribution of gene frequencies. *Proceedings of the National Academy of Sciences of the United States of America, 31*, 382–389.

Wright, S. (1948). On the roles of directed and random changes in gene frequency in the genetics of populations. *Evolution, 2*, 279–294.

Wright, S. (1986 [1939]). Statistical genetics in relation to evolution. In W. B. Provine (Ed.). *Evolution: Selected papers* (pp. 282–341). Chicago: University of Chicago Press.