

A Wright–Fisher  
model with  
indirect selection.

Ludovic  
Goudenège\* and  
Pierre-André Zitt†.

Introduction:  
models and main  
results

The  
single-generation  
model — large  
population limit

The  
multi-generation  
model

Conclusion

## A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†.

June 2, 2015

---

\*CNRS, Fédération de mathématiques de l'École Centrale Paris

†Université Paris-Est Marne-la-Vallée, LAMA

## Abstract

Dans cet exposé, je présenterai une généralisation du modèle de Wright-Fisher dans laquelle certains individus adoptent un comportement qui est délétère vis-à-vis des autres individus et sans aucun avantage direct pour eux-mêmes.

Ce comportement paradoxal est adopté par plusieurs espèces de parasites hyménoptères pour lesquelles les individus mâles continuent de s'accoupler avec des femelles alors qu'ils n'ont plus la capacité de les féconder.

Dans la limite de grande population et pour un ratio mâle/femelle fixé, on montrera la convergence du processus de reproduction renormalisé vers un processus de diffusion avec une dérive génétique et une sélection dépendantes de la fréquence.

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

Harmful behaviours and population genetics  
Basic model  
Large population limit  
Multiple generations: the classical Wright–Fisher model with selection  
A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

## Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

## The single-generation model — large population limit

Outline of the proof

Identifying the limit function

Resolution of the PDE

## The multi-generation model

Main line of the proof

The martingale problem is well posed

The number of reproductions

Infinitesimal mean and variance

Proofs for the full model

## Conclusion

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

## Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

## The single-generation model — large population limit

Outline of the proof

Identifying the limit function

Resolution of the PDE

## The multi-generation model

Main line of the proof

The martingale problem is well posed

The number of reproductions

Infinitesimal mean and variance

Proofs for the full model

## Conclusion

# Harmful behaviours and population genetics

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

The object of population genetics is to understand how the genetic composition of a population changes through time in response to mutation, natural selection and demographic stochasticity (“genetic drift”) and mutations.

In the simplest case, consider a gene with a haploid locus segregating two alleles (say “white” and “black”), which affect an individual’s phenotype. Here we are interested in the changes of the proportion of individuals carrying the white allele over generations.

One of the simplest stochastic models for this evolution is the classical Wright–Fisher model for genetic drift\*.

---

\* Let us recall the unfortunate polysemy of the word “drift”. In the biological literature “genetic drift” corresponds to the noise-induced variations. When using a stochastic model, this is at odds with the “drift” of a diffusion, i.e. the first order term that models a deterministic force.

# Harmful behaviours and population genetics

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical

Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

It is simple enough that a very detailed mathematical analysis can be performed (see for example the monographs [8] or [9], where many other questions and models are studied from a mathematical point of view).

Many variations of this model have been studied, adding selection and mutation to the picture. Classically, selection has been added to this model by stipulating that one of the alleles is  $(1 + \beta)$  more likely to be chosen for the next generation than the other allele. This models a direct advantage: for example, the eggs carrying the white allele may have more chance to mature.

## Harmful behaviours and population genetics

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical

Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

In many biological settings, individuals perform actions that may harm others without giving the perpetrator any direct advantage. For example, males of several invertebrate species have a limited sperm stock. Surprisingly, they have been reported to continue to attempt mating with virgin females while being completely sperm depleted.

Obviously, this behaviour does not aim to fertilize the eggs of these virgin females. However, in these species, copulation (with or without sperm release) has the property of stopping female sexual receptivity. This, for instance can occur as a behavioural response of the female or as a consequence of toxic seminal fluids or plugs inserted in female genitalia by males. Males can also guard the female during her receptivity period without copulating with her. These male behaviours do not increase the absolute number of eggs they fertilize.

However, because these actions limit the ability of other males to fertilize eggs, it has been suggested that they may have evolved as a male mating strategy to increase the relative number of offspring sired by individuals that use this strategy. This model is inspired by a model for the evolution of spiteful behaviour.

# Harmful behaviours and population genetics

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical

Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

Our aim is to analyze a variation of the Wright–Fisher model where such an effect appears. Quite interestingly, this model will prove to be equivalent (in the large population limit) to a model with frequency dependent selection.

The rescaled process follows the diffusion on  $[0, 1]$  given by the SDE:

$dX_t = \sqrt{a(X_t)}dB_t + b(X_t)dt$  and the corresponding generator

$L = \frac{1}{2}a(x)\partial_{xx} + b(x)\partial_x$ , where

$$\begin{cases} a(x) = \frac{x(1-x)}{v_s(x)}, \\ b(x) = x(1-x) \left( \beta - \frac{v'_s(x)}{v_s^2(x)} \right). \end{cases}$$



A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

## Introduction: models and main results

Harmful behaviours and population genetics

### Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

## The single-generation model — large population limit

Outline of the proof

Identifying the limit function

Resolution of the PDE

## The multi-generation model

Main line of the proof

The martingale problem is well posed

The number of reproductions

Infinitesimal mean and variance

Proofs for the full model

## Conclusion

## Basic model

We first define a model for one generation, where a certain number of females visit a pool of males, some of which carry the black allele that codes for the “harmful” behaviour. When the number of individuals is large we can analyze precisely the reproduction probabilities for each type of individual. Finally we show how to adapt the Wright–Fisher model to our case, and state our main result, namely a diffusion limit for the renormalized multi-generation model.

In the basic model, suggested by F.-X. Dechaume-Moncharmont and M. Galipaud, consider an urn with  $w$  white balls and  $b$  black balls. All balls begin as “unmarked”. Draw  $f$  times from this urn, with the following rule:

- ▶ if the ball drawn is white, mark it and remove it from the urn;
- ▶ if it is black and unmarked, mark it and put it back in the urn;
- ▶ if it is black and already marked, put it back in the urn.

After the  $f$  draws, call  $X$  the number of marked white balls and  $Y$  the number of marked black balls.

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

## Basic model

This models a reproductive season. The balls represent males, and each draw corresponds to a reproduction attempt by a different female. The marks represent a successful reproduction. The white balls “play fair”: if they are chosen by a female, they reproduce and retire from the game. The black balls, even after reproduction, “stay in the game”: they may be chosen again in subsequent draws.

Even if it is chosen multiple times, a black ball only reproduces once, so that black balls do not get a direct reproductive advantage from their behaviour. In particular, if the colors of all the other balls are fixed, the probability of reproduction does not depend on the ball’s color.

However, the black balls “harm” all the other balls, possibly depriving them of reproduction attempts. The variables  $X$  and  $Y$  count the number of white/black males that have reproduced.

# Basic model

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical

Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

## Remark (Simplification)

*This model is of course very simplified. In particular the mating phenotype of the males only depends on a haploid locus which is paternally inherited; this assumption does not hold for example for male hymenopteran parasitoids, who inherit their genomes from their mothers. For simplicity we restrict ourselves to one model, keeping in mind that other models may lead to different expressions of the drift and variance for the diffusion limit.*

## Basic model

To compare the two strategies, we begin by comparing two individuals. In an urn with  $w$  white balls and  $b$  black balls, we look at one particular white ball (Walt) and one particular black ball (Bob). Define the probabilities of successful reproduction by:

$$p_w(w, b, f) = \mathbb{P}[\text{Walt is chosen at least once in the } f \text{ draws}]$$

$$p_b(w, b, f) = \mathbb{P}[\text{Bob is chosen at least once in the } f \text{ draws}]$$

### Theorem

*The “harmful” males have a fitness advantage, in the sense that:*

$$p_b(w, b, f) \geq p_w(w, b, f).$$

*The inequality is strict if  $f \geq 2$  and  $w, b \geq 1$ .*

A Wright–Fisher  
model with  
indirect selection.

Ludovic  
Goudenège\* and  
Pierre-André Zitt†

Introduction:  
models and main  
results

Harmful  
behaviours and  
population  
genetics

Basic model

Large population  
limit

Multiple  
generations: the  
classical  
Wright–Fisher  
model with  
selection

A Wright–Fisher  
model with  
indirect selection

The  
single-generation  
model — large  
population limit

The  
multi-generation  
model

Conclusion

## Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

## The single-generation model — large population limit

Outline of the proof

Identifying the limit function

Resolution of the PDE

## The multi-generation model

Main line of the proof

The martingale problem is well posed

The number of reproductions

Infinitesimal mean and variance

Proofs for the full model

## Conclusion

# Large population limit

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical

Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

To quantify the advantage given by the “harmful” behaviour, it is natural to look at a large population limit, when the number of black balls (“harmful” males), white balls (regular males) and the number of draws ( $f$  i.e. females) go to infinity, while the respective proportions converge.

We can describe the limiting behaviour of  $p_b$  and  $p_w$ , and more importantly of the difference  $p_b - p_w$ , in terms of the solution  $v$  of a specific PDE.

## Large population limit

To define this function  $v$  and state the approximation result we need additional notation. The numbers of individuals  $(w, b, f)$  will correspond in the continuous limit to proportions  $(x, y, z)$  in the set:

$$\Omega = \{(x, y, z) \in \mathbb{R}_+ : x + y + z \leq 1\}.$$

For  $(x, y, z) \in \Omega$ , with  $y > 0$ , we will prove that the equation

$$x(1 - e^{-t}) + yt = z \tag{1}$$

has a unique solution  $T(x, y, z) \in (0, \infty)$ .

Define two functions  $u$  and  $v$  on  $\Omega$  by:

$$u(x, y, z) = \exp(-T(x, y, z)), \quad v(x, y, z) = 1 - u(x, y, z).$$



A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical

Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

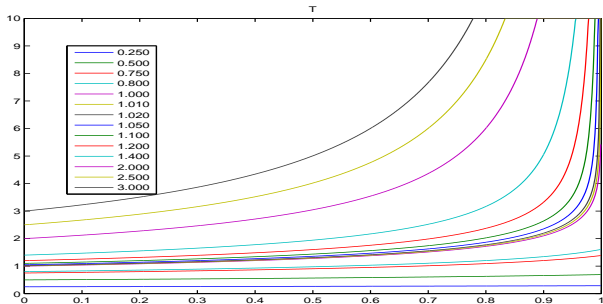


Figure: Function  $T$  in the Wright–Fisher model with indirect selection.

# A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

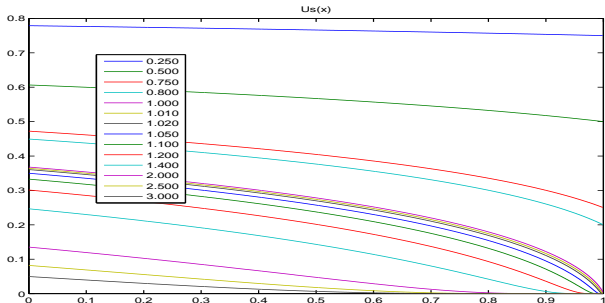


Figure: Function  $u$  in the Wright–Fisher model with indirect selection.

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

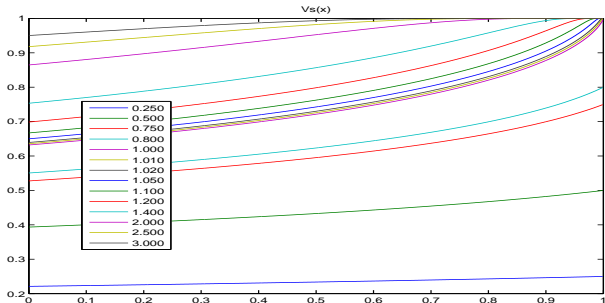


Figure: Function  $v$  in the Wright–Fisher model with indirect selection.

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†.

Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical

Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

## Large population limit

For any “population size”  $N$  we will consider functions defined on the following discretization of  $\Omega$  :

$$\Omega_N = \{(w, b, f) \in \mathbb{N}^3 : w + b + f \leq N\}.$$

For any function  $g : \Omega \rightarrow \mathbb{R}$ , we denote by  $g^N$  the discretization

$$g^N : \Omega_N \rightarrow \mathbb{R}$$

$$(w, b, f) \mapsto g\left(\frac{w}{N}, \frac{b}{N}, \frac{f}{N}\right).$$

If  $p$  is a function on  $\Omega_N$ , we denote by  $\delta_x p$ ,  $\delta_y p$  the discrete differences:

$$\delta_x p(w, b, f) = p(w + 1, b, f) - p(w, b, f),$$

$$\delta_y p(w, b, f) = p(w, b + 1, f) - p(w, b, f).$$

## Large population limit

### Remark (Technical issue)

*Most of the bounds we prove are uniform on specific subsets of  $\Omega$  or  $\Omega_N$ . For any  $y_0 > 0$ , and any  $s < 1$ , we define:*

$$\Omega(y_0) = \{(x, y, z) \in \Omega : y \geq y_0\};$$

$$\Omega_N(y_0) = \left\{ (w, b, f) \in \mathbb{N}^3 : \left( \frac{w}{N}, \frac{b}{N}, \frac{f}{N} \right) \in \Omega(y_0) \right\};$$

$$\Omega(s) = \{(x, y, z) \in \Omega : z \leq s(x + y) \text{ and } x - z \geq (1 - s)/(2 + 2s)\};$$

$$\Omega_N(s) = \left\{ (w, b, f) \in \mathbb{N}^3 : \left( \frac{w}{N}, \frac{b}{N}, \frac{f}{N} \right) \in \Omega(s) \right\}.$$

A Wright–Fisher  
model with  
indirect selection.

Ludovic  
Goudenège\* and  
Pierre-André Zitt†.

Introduction:  
models and main  
results

Harmful  
behaviours and  
population  
genetics

Basic model

Large population  
limit

Multiple  
generations: the  
classical  
Wright–Fisher  
model with  
selection

A Wright–Fisher  
model with  
indirect selection

The  
single-generation  
model — large  
population limit

The  
multi-generation  
model

Conclusion

## Large population limit

### Remark (On the sets $\Omega(y_0)$ and $\Omega(s)$ )

*Let us repeat that  $x$ ,  $y$  and  $z$  are the continuous analogues of  $w$ ,  $b$  and  $f$ . In this light,  $y_0$  corresponds to a minimal proportion of “harmful” males, and  $s$  to a maximal sex ratio.*

*The second condition appearing in the definition of  $\Omega(s)$  is less natural: it is a way of ruling out degenerate points where both  $x$  and  $y$  are small, which will be crucial for finding good bounds on  $u$ ,  $v$  and their derivatives, while keeping an essential “stability” property.*

## Large population limit

Now we can state the first asymptotic result.

### Theorem

For any  $y_0 > 0$ , there exists a constant  $C(y_0)$  such that for all  $N$ ,

$$\forall (w, b, f) \in \Omega_N(y_0), \quad \begin{aligned} \left| p_w(w, b, f) - v^N(w, b, f) \right| &\leq \frac{C(y_0)}{N}, \\ \left| p_b(w, b, f) - v^N(w, b, f) \right| &\leq \frac{C(y_0)}{N}, \end{aligned}$$

where  $v^N$  is the discretization of  $v$ . Moreover, the difference of fitness is of order  $1/N$ , and more precisely:

$$\forall (w, b, f) \in \Omega_N(y_0), \quad \left| p_b(w, b, f) - p_w(w, b, f) - \frac{1}{N}(\partial_x v - \partial_y v)^N(w, b, f) \right| \leq \frac{C(y_0)}{N^2}.$$

For any  $s < 1$ , the same bounds hold uniformly on all  $\Omega_N(s)$ , with  $C(y_0)$  replaced by a constant  $C(s)$  that only depends on  $s$ .

A Wright–Fisher  
model with  
indirect selection.

Ludovic  
Goudenège\* and  
Pierre-André Zitt†

Introduction:  
models and main  
results

Harmful  
behaviours and  
population  
genetics

Basic model

Large population  
limit

Multiple  
generations: the  
classical  
Wright–Fisher  
model with  
selection

A Wright–Fisher  
model with  
indirect selection

The  
single-generation  
model — large  
population limit

The  
multi-generation  
model

Conclusion

## Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

## The single-generation model — large population limit

Outline of the proof

Identifying the limit function

Resolution of the PDE

## The multi-generation model

Main line of the proof

The martingale problem is well posed

The number of reproductions

Infinitesimal mean and variance

Proofs for the full model

## Conclusion



# The classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

Harmful behaviours and population genetics

Basic model  
Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

The Wright–Fisher model with selection is a Markov chain  $(X_k^N)_{k \in \mathbb{N}}$  on  $\{0, 1/N, 2/N, \dots, 1\}$  that describes (a simplification of) the evolution of the frequency of an allele in a population across generations. This is a very simplified model, where the size  $N$  of the population is fixed.

To simplify the exposition suppose that the first allele is “white” and the second “black”; at time  $k$  a proportion  $X_k^N$  of the population is “white”. Given the state  $x$  at time  $k$ , the next state is chosen in the following way.

# The classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

**First step.** All individuals lay a very large number  $M$  of eggs. A proportion  $s_b(N)$  (resp.  $s_w(N)$ ) of black (resp. white) eggs survive this first step, so there are  $M \cdot N(1 - x) \cdot s_b(N)$  black eggs and  $M \cdot Nx \cdot s_w(N)$  white ones.

**Second step.** The population at time  $k + 1$ , of size  $N$ , is chosen by picking randomly  $N$  eggs among the surviving ones. Since  $M$  is very large, the number of white individuals at time  $k + 1$  is approximately binomial. If the ratio of the surviving probabilities is

$$1 + \beta(N) = s_w(N)/s_b(N),$$

then the parameters of the binomial are  $N$  and  $\frac{(1 + \beta(N))x}{(1 - x) + (1 + \beta(N))x}$ .

# The classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

In the large population limit  $N \rightarrow +\infty$ , at long time scales and in the regime of weak selection where  $\beta(N) = \beta/N$ , it is well-known that the finite size model can be approximated by a solution of a stochastic differential equation (namely a diffusion).

This use of diffusion approximations in population genetics is now well established.

More precisely, define for all  $N$  a continuous time process  $(X_t^N)_{t \geq 0}$  by:

$$\forall t \in [0, 1], X_t^N = X_{\lfloor t/N \rfloor}^N.$$

The diffusion approximation is the following:

## The classical Wright–Fisher model with selection

### Theorem (Wright–Fisher diffusion with selection)

*In the weak selection limit, the rescaled Wright–Fisher model  $(X^N)_{t \geq 0}$  converges weakly (in the Skorokhod sense) as  $N \rightarrow \infty$  to the diffusion*

$$dX_t = \sqrt{a(X_t)} dB_t + b(X_t) dt \text{ generated by } L = \frac{1}{2} a(x) \partial_{xx} + b(x) \partial_x,$$

where

$$\begin{cases} a(x) = x(1-x) \\ b(x) = \beta x(1-x). \end{cases}$$

### Remark

*If the white eggs survive better than the black ones, then  $s_w(N) > s_b(N)$  so  $\beta$  is positive; the diffusion drifts towards  $x = 1$ . If black eggs are favored,  $\beta$  is negative and the drift is towards 0.*

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

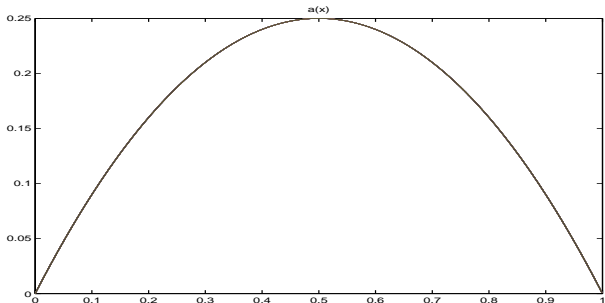


Figure: Function  $a$  in the Wright–Fisher model with selection.

# A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical

Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

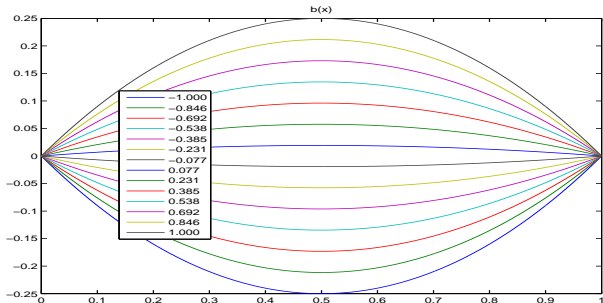


Figure: Function  $b$  in the Wright–Fisher model with selection.

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

## Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

### The single-generation model — large population limit

Outline of the proof

Identifying the limit function

Resolution of the PDE

### The multi-generation model

Main line of the proof

The martingale problem is well posed

The number of reproductions

Infinitesimal mean and variance

Proofs for the full model

## Conclusion

## A Wright–Fisher model with indirect selection

Let us now see how the basic model can be used as a building block for a multiple generation model in the spirit of the classical Wright–Fisher model, in order to study the evolution of the “harmful” trait along generations.

In the literature, various extensions of the Wright–Fisher model have been considered under various scalings.

Frequency-dependent coefficients may appear in such models but are often built-in in the individual-based model. Another possible extension is to make the offspring play a (game-theoretic) game after the random mating step.

Frequency dependence appears more naturally for modelling resistance to epidemics and comparing offspring distributions with different variances; these papers do not however link the diffusion to a precise individual-based model.

In the more complicated setting of the evolution of continuous traits, several papers start with individual-based models and establish rigorously various limits, showing convergence to deterministic processes, SDEs or solutions of integro-differential equations. Finally in the literature a popular study concerns the fixation probabilities and problems arising at the boundaries.



## A Wright–Fisher model with indirect selection

Here we fix once and for all a sex-ratio by fixing the parameter  $s > 0$ , and supposing that there are  $s$  females for one male, *i.e.* a proportion  $1/(1+s) \in (0, 1)$  of the total population is male.

Consider a large urn with  $n$  (male) balls, let  $f_n = \lfloor sn \rfloor$ , and define the state space  $\mathcal{S}_n = \left\{ 0, \frac{1}{n}, \frac{2}{n}, \dots, 1 \right\}$ : these are the possible values for the proportion of white balls.

We define an  $\mathcal{S}_n$ -valued Markov chain  $(X_k^n)_{k \in \mathbb{N}}$  as follows. Suppose that the initial proportion of white balls at time  $k = 0$  in the urn is  $X_0^n = x \in \mathcal{S}_n$ : there are  $w = xn$  white balls and  $b = (1-x)n$  black balls. The next state  $X_1^n$  is chosen in two steps.

## A Wright–Fisher model with indirect selection

**First step.** The  $f_n$  female pick partners according to the single-generation model introduced previously: this leads to  $\tilde{X}_1^n$  reproduction with normal males and  $\tilde{Y}_1^n$  reproduction with “harmful” males. As before, each of these reproductions creates a very large number of “eggs”.

A proportion  $s_w(N)$  (resp.  $s_b(N)$ ) of white (resp. black) eggs survive, and the ratio  $s_w(N)/s_b(N)$  is still denoted by  $1 + \beta(N)$  with  $\beta(N) = \beta/N$ .

After this step there is a very large number of eggs, a proportion

$$\tilde{Z}_{1,\beta}^n = \frac{(1 + \beta(N))\tilde{X}_1^n}{(1 + \beta(N))\tilde{X}_1^n + \tilde{Y}_1^n}$$

of which are white.

**Second step.** Among all the eggs,  $n$  eggs are chosen uniformly at random. Once more, since the number of eggs is supposed to be very large, the number of white balls in the next generation follows a binomial law of parameters  $n$  and  $\tilde{Z}_{1,\beta}^n$ . Finally divide this number by  $n$  to get  $X_1^n$ , the proportion of white balls at time  $k = 1$ .

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical

Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

# A Wright–Fisher model with indirect selection

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical

Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

We iterate the process to define  $(X_k^n)_{k \geq 2}$ . As above we define a continuous process by accelerating time and let:

$$\forall t \geq 0, X_t^n = X_{\lfloor t/n \rfloor}^n.$$

Our main result is a diffusion limit for the rescaled process  $\left(\frac{1}{n}X_k^n\right)_k$  with an explicit non-trivial drift towards 0. The drift and volatility are expressed in terms of the following function:

$$v_s : [0, 1] \rightarrow \mathbb{R},$$

$$x \mapsto v \left( \frac{x}{1+s}, \frac{1-x}{1+s}, \frac{s}{1+s} \right),$$

where we recall that  $v$  is defined by  $1 - \exp(-T)$ .

## A Wright–Fisher model with indirect selection

### Theorem

If  $s < 1$ , the rescaled process  $X_t^n$  converges weakly (in the Skorokhod sense) to the diffusion on  $[0, 1]$  given by the SDE:

$$dX_t = \sqrt{a(X_t)}dB_t + b(X_t)dt$$

and the corresponding generator

$$L = \frac{1}{2}a(x)\partial_{xx} + b(x)\partial_x,$$

where

$$\begin{cases} a(x) = \frac{x(1-x)}{v_s(x)}, \\ b(x) = x(1-x) \left( \beta - \frac{v_s'(x)}{v_s^2(x)} \right). \end{cases}$$

# A Wright–Fisher model with indirect selection

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical

Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

## Remark

*If  $s \geq 1$ , we are only able to prove the convergence until the process reaches  $x = 1 - y_0$ ; we currently do not know whether or not the behaviours at the boundary  $x = 1$  differ for the discrete and continuous process.*

*This possibly purely technical restriction prevents us from rigorously justifying the approximation of the discrete absorption probabilities and mean absorption time by their continuous counterparts, which is one of the usual applications for diffusion approximations.*

## Remark

*The function  $v_s$  is very nice, in particular it is strictly increasing ( $v'_s > 0$ ). If  $s < 1$ , it is bounded away from zero.*

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

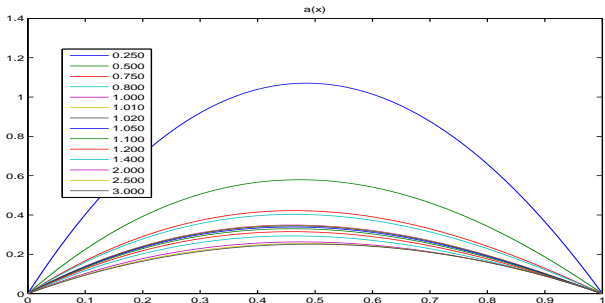


Figure: Function  $a$  in the Wright–Fisher model with indirect selection.

## A Wright–Fisher model with indirect selection

Let us just stress two points as regards the comparison with the classical model:

The variance is multiplied by  $(1/v_s(x)) > 1$ ; this is a natural consequence of the additional noise in the first step. The precise factor may be heuristically justified as follows: when  $n$  is large, there are  $nv_s(x) + \mathcal{O}(1)$  successful reproductions, thus, with binomial resampling of offspring, the male variance effective population size is also  $nv_s(x) + \mathcal{O}(1)$ .

To compare the drift coefficients, it is natural to consider the “normalized” quantity  $2b = a$  which fully determines the scale functions and hitting probabilities. In this light, up to a change of time, our modified diffusion corresponds to the classical one with a selection parameter  $\beta(x) = \beta v_s(x) - \frac{v_s'(x)}{v_s(x)}$  that depends on  $x$ . If  $s \rightarrow \infty$  this goes to  $\beta$ : all males have a chance to reproduce and the harmful strategy has no effect. If  $\beta = 0$ ,  $\beta(x)$  is negative (and there is a non trivial drift towards 0). In the general case, depending on the values of  $\beta$  and  $s$ , there may be one or more “equilibrium” points where the drift cancels out.

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

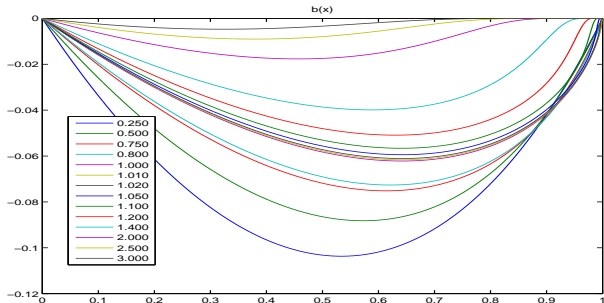


Figure: Function  $b$  in the Wright–Fisher model with indirect selection.  $\beta = 0$



## A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

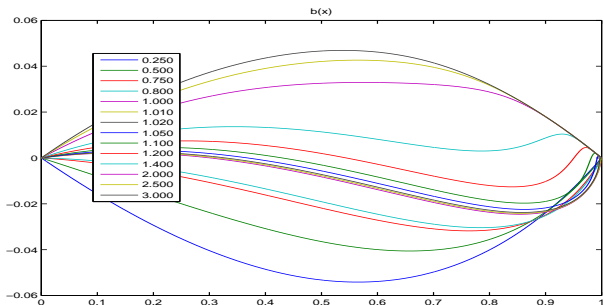


Figure: Function  $b$  in the Wright–Fisher model with indirect selection.  $\beta = 0.2$

# A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical

Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

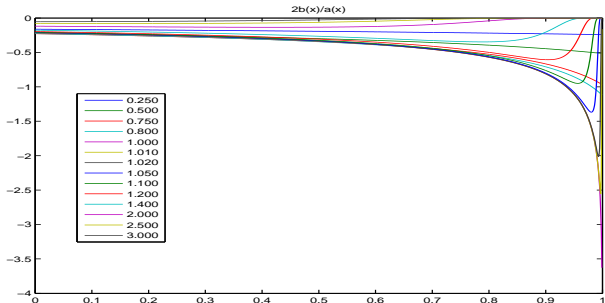


Figure: Function  $2b/a$  in the Wright–Fisher model with indirect selection.  $\beta = 0$

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

The single-generation model — large population limit

The multi-generation model

Conclusion

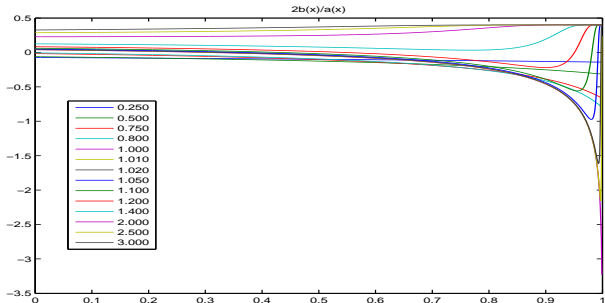


Figure: Function  $2b/a$  in the Wright–Fisher model with indirect selection.  $\beta = 0.2$

A Wright–Fisher  
model with  
indirect selection.

Ludovic  
Goudenège\* and  
Pierre-André Zitt†

Introduction:  
models and main  
results

The  
single-generation  
model — large  
population limit

Outline of the  
proof  
Identifying the  
limit function  
Resolution of the  
PDE

The  
multi-generation  
model

Conclusion

## Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

## The single-generation model — large population limit

Outline of the proof

Identifying the limit function

Resolution of the PDE

## The multi-generation model

Main line of the proof

The martingale problem is well posed

The number of reproductions

Infinitesimal mean and variance

Proofs for the full model

## Conclusion

A Wright–Fisher  
model with  
indirect selection.

Ludovic  
Goudenège\* and  
Pierre-André Zitt†

Introduction:  
models and main  
results

The  
single-generation  
model — large  
population limit

Outline of the  
proof

Identifying the  
limit function

Resolution of the  
PDE

The  
multi-generation  
model

Conclusion

## Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

## The single-generation model — large population limit

### Outline of the proof

Identifying the limit function

Resolution of the PDE

## The multi-generation model

Main line of the proof

The martingale problem is well posed

The number of reproductions

Infinitesimal mean and variance

Proofs for the full model

## Conclusion

## Outline of the proof

The goal of this section is to prove the theorem on the convergence of  $p_b$  and  $p_w$  to a continuous function  $v$ .

Add a single red ball to the  $w$  white and  $b$  black balls; let us call it Roger. Draw from the urn until the red ball is drawn or we have made  $f$  draws; the white balls are not replaced but the black ones are. Define:

$$q(w, b, f) = \mathbb{P}[\text{Roger is not drawn}].$$

Since the color of a ball only matters if it is drawn, and only influences the subsequent draws, it is easy to see that:

$$1 - p_w(w, b, f) = q(w - 1, b, f), \quad 1 - p_b(w, b, f) = q(w, b - 1, f).$$

Therefore it is enough to compare the probabilities that the red ball is never drawn, when one ball goes from black to white.

## Outline of the proof

It will be slightly easier to work on the quantity  $q(w, b, f)$ , and deduce the statements on  $p_w$  and  $p_b$  afterwards.

This discrete function  $q$  approximates the function  $u$ , and we also get convergence of the discrete differences of  $q$  to the derivatives of  $u$ :

### Theorem

*For all  $y_0 > 0$ , there exists  $C(y_0)$  such that, for all  $N$ , and all  $(w, b, f) \in \Omega_N(y_0)$ ,*

$$\left| (q - u^N)(w, b, f) \right| \leq \frac{C(y_0)}{N},$$

$$\left| (N\delta_x q - (\partial_x u)^N)(w, b, f) \right| \leq \frac{C(y_0)}{N},$$

$$\left| (N\delta_y q - (\partial_y u)^N)(w, b, f) \right| \leq \frac{C(y_0)}{N}.$$

*If  $s < 1$ , there exists  $C(s)$  such that the same bounds hold uniformly on  $\Omega_N(s)$ , where  $C(y_0)$  is replaced by  $C(s)$  on the right hand side.*

## Outline of the proof

The proof hinges on the following recurrence relation for  $q$ , which follows by conditioning on the result of the first draw:

$$q(w, b, f) = \frac{w}{w + b + 1} q(w - 1, b, f - 1) + \frac{b}{w + b + 1} q(w, b, f - 1)$$

The main idea is then to view  $q$  as a discrete version of  $u$ , and the recurrence relation as an approximation of a relation between derivatives of  $u$ . The corresponding PDE for  $u$  is explicitly solvable.



A Wright–Fisher  
model with  
indirect selection.

Ludovic  
Goudenège\* and  
Pierre-André Zitt†

Introduction:  
models and main  
results

The  
single-generation  
model — large  
population limit  
Outline of the  
proof

Identifying the  
limit function

Resolution of the  
PDE

The  
multi-generation  
model

Conclusion

## Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

## The single-generation model — large population limit

Outline of the proof

Identifying the limit function

Resolution of the PDE

## The multi-generation model

Main line of the proof

The martingale problem is well posed

The number of reproductions

Infinitesimal mean and variance

Proofs for the full model

## Conclusion

## Identifying the limit function

Let us now give a short heuristic argument for finding the limit function  $u$ . Suppose that  $u$  exists, and that all of the limits encountered below converge. Starting from a recurrence relation on  $q$ , we introduce  $q(w, b, f)$  on the right hand side, so that discrete differences appear:

$$\begin{aligned}q(w, b, f) &= \frac{w}{w + b + 1} q(w - 1, b, f - 1) + \frac{b}{w + b + 1} q(w, b, f - 1) \\&= \frac{w + b}{w + b + 1} q(w, b, f) \\&\quad + \frac{w}{w + b + 1} (q(w - 1, b, f - 1) - q(w, b, f)) \\&\quad + \frac{b}{w + b + 1} (q(w, b, f - 1) - q(w, b, f)).\end{aligned}$$

Multiplying by  $(w + b + 1)$ , we find after simplification:

$$\begin{aligned}q(w, b, f) &= w(q(w - 1, b, f - 1) - q(w, b, f)) \\&\quad + b(q(w, b, f - 1) - q(w, b, f)) \\&= (w/N) \cdot N(q(w - 1, b, f - 1) - q(w, b, f)) \\&\quad + (b/N) \cdot N(q(w, b, f - 1) - q(w, b, f)).\end{aligned}$$

## Identifying the limit function

Now if  $N = w + b + f$  goes to  $\infty$ , and if  $(w/N, b/N, f/N)$  converges to  $(x, y, z)$ , the left hand side converges to  $u$  and the right hand side to  $x(-(\partial_x + \partial_z)u) + y(-\partial_z u)$ , so that  $u$  satisfies

$$u + x\partial_x u + (x + y)\partial_z u = 0.$$

Since  $q(w, b, 0) = 1$ , we also obtain  $u(x, y, 0) = 1$ . Summing up, if  $q(w, b, f)$  converges “in a good way” to a function  $u$ , this function satisfies an explicit first-order PDE on  $\Omega$ :

$$\begin{cases} \forall (x, y, z) \in \Omega, & u + F \cdot \nabla u = 0, \\ \forall (x, y), & u(x, y, 0) = 1, \end{cases}$$

where  $F$  is the vector field  $F(x, y, z) = (x, 0, x + y)$ .

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

The single-generation model — large population limit  
Outline of the proof  
Identifying the limit function

Resolution of the PDE

The multi-generation model

Conclusion

## Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

## The single-generation model — large population limit

Outline of the proof

Identifying the limit function

Resolution of the PDE

## The multi-generation model

Main line of the proof

The martingale problem is well posed

The number of reproductions

Infinitesimal mean and variance

Proofs for the full model

## Conclusion

## Resolution of the PDE

This first order PDE can be solved by the method of characteristics. We look for trajectories  $M(t) = (x(t); y(t); z(t))$  that satisfy the characteristic equation:

$$\frac{d}{dt}M(t) = -F(M(t))$$

The solution is:

$$\begin{cases} x(t) = x_0 e^{-t} \\ y(t) = y_0 \\ z(t) = x_0(e^{-t} - 1) - y_0 t + z_0. \end{cases}$$

Now  $h(t) = u(M(t))$  satisfies:

$$\frac{dh}{dt} = \nabla u \cdot \frac{d}{dt}M(t) = -\nabla_u(M(t)) \cdot F(M(t)) = u(M(t)) = h(t).$$

Therefore  $h(t) = h(0) \exp(t)$ .

Suppose  $T = T(x_0, y_0, z_0)$  is a solution of the third equation, i.e.  $z(T) = 0$ . Then  $h(T) = u(M(T)) = 1$  thanks to the boundary condition.

Finally :

$$u(x_0, y_0, z_0) = h(0) = u(M(T)) \exp(-T) = \exp(-T(x_0, y_0, z_0)).$$

## Resolution of the PDE

### Theorem (Properties of the solution)

*If  $(x, y, z) \in \Omega$  and if  $y > 0$ , the equation (1) defining  $T$  has a unique solution. The function  $u$  is smooth on the interior domain  $\{(x, y, z) \in (\mathbb{R}_+^*)^3, x + y + z < 1\}$ . For any  $y_0 > 0$ , there exists a constant  $C(y_0)$  such that for all  $(x, y, z) \in \Omega(y_0)$ , and all  $(i, j)$ ,*

$$|u(x, y, z)| \leq C(y_0), \quad |\partial_i u(x, y, z)| \leq C(y_0), \quad |\partial_i \partial_j u(x, y, z)| \leq C(y_0).$$

*If  $s < 1$ , similar bounds hold uniformly on  $\Omega(s)$ .*

A Wright–Fisher  
model with  
indirect selection.

Ludovic  
Goudenège\* and  
Pierre-André Zitt†

Introduction:  
models and main  
results

The  
single-generation  
model — large  
population limit

The  
multi-generation  
model

Main line of the  
proof  
The martingale  
problem is well  
posed  
The number of  
reproductions  
Infinitesimal mean  
and variance  
Proofs for the full  
model

Conclusion

## Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

## The single-generation model — large population limit

Outline of the proof

Identifying the limit function

Resolution of the PDE

## The multi-generation model

Main line of the proof

The martingale problem is well posed

The number of reproductions

Infinitesimal mean and variance

Proofs for the full model

## Conclusion

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

The single-generation model — large population limit

The multi-generation model

Main line of the proof

The martingale problem is well posed

The number of reproductions

Infinitesimal mean and variance

Proofs for the full model

Conclusion

## Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

## The single-generation model — large population limit

Outline of the proof

Identifying the limit function

Resolution of the PDE

## The multi-generation model

Main line of the proof

The martingale problem is well posed

The number of reproductions

Infinitesimal mean and variance

Proofs for the full model

## Conclusion



## The multi-generation model : Main line of the proof

To prove the diffusion limit, we follow the presentation of Durrett in [7].

For each  $n$ , we have defined a Markov chain  $(X_k^n)_{k \in \mathbb{N}}$ , that lives on the state space  $\mathcal{S}_n = \{0, \frac{1}{n}, \dots, 1\} \subset \mathbb{R}$ .

Let  $\mathbb{E}_x[\cdot]$  and  $\mathbf{Var}_x(\cdot)$  denote the expectation and variance operators for the Markov chain started at  $X_0^n = x$ . Define, for each  $n$  and each  $x \in \mathcal{S}_n$ , the “infinitesimal variance”  $a^n(x)$  and the “infinitesimal mean”  $b^n(x)$  by:

$$a^n(x) = n \mathbf{Var}_x(X_1^n),$$

$$b^n(x) = n(\mathbb{E}_x[X_1^n] - x),$$

and let

$$c^n(x) = n \mathbb{E}_x[|X_1^n - x|^3].$$

Suppose additionally that  $a$  and  $b$  are two continuous functions for which the martingale problem is well posed, i.e., for each  $x$  there is a unique measure  $P_x$  on  $\mathcal{C}([0, \infty), \mathbb{R})$  such that  $P_x[X_0 = x] = 1$  and

$$X_t - \int_0^t b(X_s) ds \quad \text{and} \quad X_t^2 - \int_0^t a(X_s) ds$$

are local martingales. In this setting, the convergence of the discrete process to its limit is a consequence of the following result.

## Main line of the proof

### Theorem

Suppose that the following three conditions hold.

1. The infinitesimal mean and variance converge uniformly:

$$\limsup_n \sup_{x \in \mathcal{S}_n} |a^n(x) - a(x)| = 0,$$

$$\limsup_n \sup_{x \in \mathcal{S}_n} |b^n(x) - b(x)| = 0.$$

2. The size of the discrete jumps is small enough:

$$\limsup_n \sup_{x \in \mathcal{S}_n} c^n(x) = 0.$$

3. The initial condition  $X_0^n = x^n$  converges to  $x$ .

Then the renormalized process converges to the diffusion  $X_t$ .

### Remark

The original formulation is  $d$ -dimensional and considers diffusions on the whole space, therefore it includes additional details that will not be needed here.

Using this result, the theorem will follow once we prove that the martingale problem is well posed and we show the following estimates.

## Main line of the proof

### Proposition (Infinitesimal mean and variance)

The following estimates hold:

$$a_n(x) = \frac{x(1-x)}{v_s(x)} + \mathcal{O}(1/\sqrt{n}),$$

$$b_n(x) = x(1-x) \left( \beta - \frac{v'_s(x)}{v_s^2(x)} \right) + \mathcal{O}(1/\sqrt{n}),$$

$$c_n(x) = \mathcal{O}(1/\sqrt{n}),$$

where the “ $\mathcal{O}$ ” holds:

- ▶ uniformly on  $\mathcal{S}_n \cap [0, x_0]$ , for all  $x_0 < 1$ , if  $s \geq 1$ ,
- ▶ uniformly on the entire space  $\mathcal{S}_n$ , if  $s < 1$ .

A Wright–Fisher  
model with  
indirect selection.

Ludovic  
Goudenège\* and  
Pierre-André Zitt†

Introduction:  
models and main  
results

The  
single-generation  
model — large  
population limit

The  
multi-generation  
model

Main line of the  
proof

The martingale  
problem is well  
posed

The number of  
reproductions  
Infinitesimal mean  
and variance  
Proofs for the full  
model

Conclusion

## Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

## The single-generation model — large population limit

Outline of the proof

Identifying the limit function

Resolution of the PDE

## The multi-generation model

Main line of the proof

The martingale problem is well posed

The number of reproductions

Infinitesimal mean and variance

Proofs for the full model

## Conclusion

## The martingale problem is well posed

Recall the definition of the function  $v_s$ :

$$\forall x \in [0, 1), \quad v_s(x) = v\left(\frac{x}{1+s}, \frac{1-x}{1+s}, \frac{s}{1+s}\right)$$

This function is extended by continuity at point  $x = 1$  by  $v_s(1) = \min(s, 1)$ . This function behaves nicely, at least if  $s < 1$ .

### Lemma (Properties of $v_s$ )

- ▶ For all  $s \in \mathbb{R}_+$ , for all  $x \in [0, 1]$ ,  $1 - e^{-s} \leq v_s(x) \leq \min(s, 1)$ ,
- ▶ For all  $s \in \mathbb{R}_+$ , for all  $x \in [0, 1)$ ,  $v_s(x) < \min(s, 1)$  and  $v'_s(x) > 0$ ,
- ▶ For all  $s < 1$ , for all  $x \in [0, 1]$ ,

$$(1-s)(e^{-s} + s - 1) \leq v'_s(x) \leq \frac{e^{-s}(-s - \log(1-s))}{(1-s)},$$

- ▶ For all  $s < 1$ , for all  $x \in [0, 1]$ ,

$$s(1-s)^2(e^{-s} + s - 1) \leq v''_s(x) \leq \frac{2se^{-s}(-s - \log(1-s))}{(1-s)^3}.$$

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

The single-generation model — large population limit

The multi-generation model

Main line of the proof

The martingale problem is well posed

The number of reproductions  
Infinitesimal mean and variance  
Proofs for the full model

Conclusion

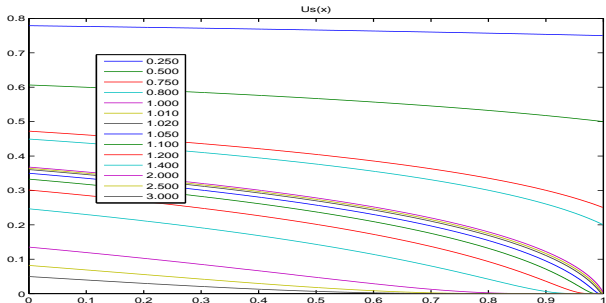


Figure: Function  $u_s$  in the Wright–Fisher model with indirect selection.

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

The single-generation model — large population limit

The multi-generation model

Main line of the proof

The martingale problem is well posed

The number of reproductions

Infinitesimal mean and variance

Proofs for the full model

Conclusion

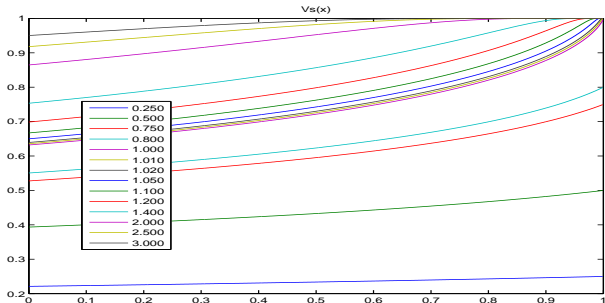


Figure: Function  $v_s$  in the Wright–Fisher model with indirect selection.

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

The single-generation model — large population limit

The multi-generation model

Main line of the proof

The martingale problem is well posed

The number of reproductions  
Infinitesimal mean and variance  
Proofs for the full model

Conclusion

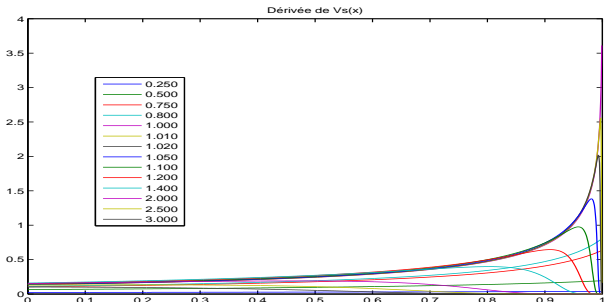


Figure: Function  $v'_s$  in the Wright–Fisher model with indirect selection.



## The martingale problem is well posed

Now we are able to prove that the martingale problem is well posed by proving pathwise uniqueness thanks to the following theorem of Yamada and Watanabe, as stated in [7, Theorem 5.3.3].

### Theorem (Yamada-Watanabe)

Let  $dX_t = \sqrt{a(x)}dB_t + b(X_t)dt$  be a SDE such that

1. there exists a positive increasing function  $\rho$  on  $(0, +\infty)$  such that

$$\left| \sqrt{a(x)} - \sqrt{a(y)} \right| \leq \rho(|x - y|), \quad \text{for all } x, y \in \mathbb{R}$$

and

$$\int_{]0,1[} \rho^{-2}(u) du = +\infty.$$

2. there exists a positive increasing concave function  $\kappa$  on  $(0, +\infty)$  such that

$$|b(x) - b(y)| \leq \kappa(|x - y|), \quad \text{for all } x, y \in \mathbb{R}$$

and

$$\int_{]0,1[} \kappa^{-1}(u) du = +\infty.$$

Then pathwise uniqueness holds for the SDE.

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

The single-generation model — large population limit

The multi-generation model

Main line of the proof  
The martingale problem is well posed

The number of reproductions

Infinitesimal mean and variance  
Proofs for the full model

Conclusion

## Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

## The single-generation model — large population limit

Outline of the proof

Identifying the limit function

Resolution of the PDE

## The multi-generation model

Main line of the proof

The martingale problem is well posed

The number of reproductions

Infinitesimal mean and variance

Proofs for the full model

## Conclusion

## The number of reproductions

We use the previous results to study the “first step” of each generation, getting information on the asymptotics of the random number of reproductions.

We only need to study what happens in one step of the Markov chain.

We let  $x = X_0^n = w/n \in \mathcal{S}_n$  be the initial proportion of white balls. There are  $b$  black balls, where  $b + w = n$  and we draw  $f = sn$  times. Note that  $N = w + b + f = (1 + s)n$ , and  $s$  is fixed, so  $n$  and  $N$  are of the same order.

We omit the “size” index  $n$  and the time index  $k = 1$ , denoting by  $(\tilde{X}, \tilde{Y}) = (\tilde{X}_1^n, \tilde{Y}_1^n)$  the number of white/black “reproductions” and by  $X = X_1^n$  the proportion of white balls after the first step. Moreover we let

$$\tilde{x} = \mathbb{E}_x [\tilde{X}], \quad \tilde{y} = \mathbb{E}_x [\tilde{Y}].$$

The goal of this section is to prove the following estimates.

## The number of reproductions

### Proposition (Moments of $(\tilde{X}, \tilde{Y})$ )

*The moments of  $(\tilde{X}, \tilde{Y})$  have the following asymptotic behaviour:*

$$\tilde{x} = \mathbb{E}_x [\tilde{X}] = nxv_s(x) + \mathcal{O}(1), \quad \tilde{y} = \mathbb{E}_x [\tilde{Y}] = n(1-x)v_s(x) + \mathcal{O}(1),$$

$$\mathbf{Var}_x (\tilde{X}) = \mathcal{O}(n), \quad \mathbf{Var}_x (\tilde{Y}) = \mathcal{O}(n),$$

$$\mathbf{Cov}_x (\tilde{X}, \tilde{Y}) = \mathcal{O}(n),$$

*In all these results the “ $\mathcal{O}$ ” are uniform on the starting point  $x \in [0, 1 - y_0]$  (if  $s \geq 1$ ), and uniform on  $x \in [0, 1]$  (if  $s < 1$ ).*

A Wright–Fisher  
model with  
indirect selection.

Ludovic  
Goudenège\* and  
Pierre-André Zitt†

Introduction:  
models and main  
results

The  
single-generation  
model — large  
population limit

The  
multi-generation  
model

Main line of the  
proof

The martingale  
problem is well  
posed

The number of  
reproductions

Infinitesimal mean  
and variance

Proofs for the full  
model

Conclusion

## Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

## The single-generation model — large population limit

Outline of the proof

Identifying the limit function

Resolution of the PDE

## The multi-generation model

Main line of the proof

The martingale problem is well posed

The number of reproductions

Infinitesimal mean and variance

Proofs for the full model

## Conclusion

## Infinitesimal mean and variance

From this moment on, the parameter  $\beta$  is fixed to 0. We set

$$\tilde{Z} = \frac{\tilde{X}}{\tilde{X} + \tilde{Y}}$$

so we have

$$\begin{aligned}\mathbb{E}_x[X] &= \mathbb{E}_x[\mathbb{E}_x[X|\mathcal{F}]] \\ &= \mathbb{E}_x[\tilde{Z}],\end{aligned}$$

And it makes sense to study the first moment of  $\tilde{Z}$ .

### Lemma (Expectation of $\tilde{Z}$ )

*The first moment of  $\tilde{Z}$  is given by*

$$\mathbb{E}[\tilde{Z}] - x = -\frac{1}{n} \cdot \frac{x(1-x)v'_s(x)}{v_s^2(x)} + \mathcal{O}(1/n^{3/2}).$$

### Lemma (Variance of $\tilde{Z}$ )

*The variance of  $\tilde{Z}$  is given by*

$$\mathbf{Var}(\tilde{Z}) = \frac{1}{n} x(1-x) \frac{v_s(x)(1-v_s(x))}{v_s(x)^2} + \mathcal{O}(1/n^{3/2}).$$

## Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

## The single-generation model — large population limit

Outline of the proof

Identifying the limit function

Resolution of the PDE

## The multi-generation model

Main line of the proof

The martingale problem is well posed

The number of reproductions

Infinitesimal mean and variance

Proofs for the full model

## Conclusion

## Proofs for the full model

Now we show how to compute the infinitesimal mean and variance in the general case, i.e. when  $\beta$  is arbitrary.

We still denote by  $\tilde{X}$ ,  $\tilde{Y}$  the number of white/black reproductions, by  $\mathcal{F} = \sigma(\tilde{X}, \tilde{Y})$  the corresponding  $\sigma$ -field, and by  $\tilde{Z}$  the “raw” ratio

$\tilde{Z} = \frac{\tilde{X}}{\tilde{X} + \tilde{Y}}$ . We define

$$\tilde{Z}_\beta = \frac{(1 + \beta/n)\tilde{X}}{(1 + \beta/n)\tilde{X} + \tilde{Y}},$$

so that, conditionally on  $\mathcal{F}$ ,  $n\tilde{X}$  follows a binomial law of parameters  $n$  and  $\tilde{Z}_\beta$ .



## Proofs for the full model

This modified ratio is not far from  $\tilde{Z}$ :

$$\begin{aligned}\tilde{Z}_\beta &= \frac{(1 + \beta/n)\tilde{X}}{(1 + \beta/n)\tilde{X} + \tilde{Y}} = \tilde{Z} \left( \frac{1 + \beta/n}{(1 + \beta/n)\tilde{Z} + 1 - \tilde{Z}} \right) \\ &= \tilde{Z}(1 + \beta/n)(1 - (\beta/n)\tilde{Z} + \mathcal{O}(1/n^2)) \\ &= \tilde{Z} \left( 1 + (\beta/n)(1 - \tilde{Z}) + \mathcal{O}(1/n^2) \right),\end{aligned}$$

where the  $\mathcal{O}$  is uniform on  $x$  and  $\omega$  since  $\tilde{Z}$  is bounded. Taking the expectation gives

$$\begin{aligned}\mathbb{E} \left[ \tilde{Z}_\beta \right] &= \mathbb{E} \left[ \tilde{Z} \right] + (\beta/n)\mathbb{E} \left[ \tilde{Z}(1 - \tilde{Z}) \right] + \mathcal{O}(1/n^2) \\ &= \mathbb{E} \left[ \tilde{Z} \right] + (\beta/n)\mathbb{E} \left[ \tilde{Z} \right] \left( 1 - \mathbb{E} \left[ \tilde{Z} \right] \right) - (\beta/n)\mathbf{Var} \left( \tilde{Z} \right) + \mathcal{O}(1/n^2).\end{aligned}$$

## Proofs for the full model

Now let us recall the results from previous lemmas:

$$\mathbb{E} \left[ \tilde{Z} \right] - x = -\frac{1}{n} \cdot \frac{x(1-x)v'_s(x)}{v_s^2(x)} + \mathcal{O}(1/n^{3/2}),$$
$$\text{Var} \left( \tilde{Z} \right) = \frac{1}{n} x(1-x) \frac{v_s(x)(1-v_s(x))}{v_s(x)^2} + \mathcal{O}(1/n^{3/2}).$$

This immediately entails

$$\mathbb{E} \left[ \tilde{Z}_\beta \right] - x = -\frac{1}{n} \cdot \frac{x(1-x)v'_s(x)}{v_s^2(x)} + x(1-x) \frac{\beta}{n} + \mathcal{O}(1/n^{3/2}),$$

and proves the general form of the infinitesimal mean announced in Proposition 1.

## Introduction: models and main results

Harmful behaviours and population genetics

Basic model

Large population limit

Multiple generations: the classical Wright–Fisher model with selection

A Wright–Fisher model with indirect selection

## The single-generation model — large population limit

Outline of the proof

Identifying the limit function

Resolution of the PDE

## The multi-generation model

Main line of the proof

The martingale problem is well posed

The number of reproductions

Infinitesimal mean and variance

Proofs for the full model

## Conclusion

This study stems from discussions with the biologists F.-X. Dechaume-Moncharmont and M. Galipaud, who came up with the single generation model for “indirect” fitness.

We thank them for stimulating discussions on the biological aspects of the problem.

On the mathematical side, our special thanks go to R. Eymard for his input on the inverse numerical analysis problem; we also thank V.-C. Tran and D. Coupier for interesting discussions.

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†.

Introduction: models and main results

The single-generation model — large population limit

The multi-generation model

Conclusion



F. A. C. C. Chalub and M. O. Souza.

A non-standard evolution problem arising in population genetics.  
*Commun. Math. Sci.*, 7(2):489–502, 2009.



F. A. C. C. Chalub and M. O. Souza.

The frequency-dependent Wright-Fisher model: diffusive and non-diffusive approximations.  
*J. Math. Biol.*, 68(5):1089–1133, 2014.



N. Champagnat, R. Ferriere, and S. Meleard.

Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models.  
*Theor Popul Biol*, 69(3):297–321, May 2006.



N. Champagnat, R. Ferrière, and S. Méléard.

From individual stochastic processes to macroscopic models in adaptive evolution.  
*Stoch. Models*, 24(suppl. 1):2–44, 2008.



D. Damiens and G. Boivin.

Why do sperm-depleted parasitoid males continue to mate?  
*Behavioral Ecology*, 17(1):138–143, Jan. 2006.



F. Dionisio.

Selfish and spiteful behaviour through parasites and pathogens.  
*Evol. Ecol. Res.*, 9(7):1199–1210, November 2007.



R. Durrett.

*Stochastic calculus: a practical introduction.*  
Probability and Stochastics Series. CRC PressINC, 1996.



R. Durrett.

*Probability models for DNA sequence evolution.*  
Probability and its Applications (New York). Springer, New York, second edition, 2008.



**A. Etheridge.**

*Some mathematical models from population genetics*, volume 2012 of *Lecture Notes in Mathematics*.  
Springer, Heidelberg, 2011.  
Lectures from the 39th Probability Summer School held in Saint-Flour, 2009.



**K. R. Foster, T. Wenseleers, and F. L. W. Ratnieks.**

**Spite: Hamilton's unproven theory.**  
*Annales Zoologici Fennici*, 38(3-4):229–238, 2001.



**J. H. Gillespie.**

**Natural selection for within-generation variance in offspring number.**  
*Genetics*, 76(3):601–606, Mar 1974.



**J. H. Gillespie.**

**Natural selection for within-generation variance in offspring number II. Discrete haploid models.**  
*Genetics*, 81(2):403–413, Oct 1975.



**W. D. Hamilton.**

**Selfish and spiteful behaviour in an evolutionary model.**  
*Nature*, 228(5277):1218–1220, Dec. 1970.



**S. Lessard.**

**Long-term stability from fixation probabilities in finite populations: New perspectives for ESS theory.**  
*Theoret. Popul. Biology*, 68(1):19–27, 2005.



**A. J. McKane and D. Waxman.**

**Singular solutions of the diffusion equation of population genetics.**  
*J. Theoret. Biol.*, 247(4):849–858, 2007.



**P. Radhakrishnan, D. Pérez-Staples, C. W. Weldon, and P. W. Taylor.**

**Multiple mating and sperm depletion in male queensland fruit flies: effects on female remating behaviour.**  
*Animal Behaviour*, 78(4):839–846, Oct. 2009.

A Wright–Fisher model with indirect selection.

Ludovic Goudenège\* and Pierre-André Zitt†

Introduction: models and main results

The single-generation model — large population limit

The multi-generation model

Conclusion



W. R. Rice.

Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. *Nature*, 381(6579):232–234, May 1996.



M. Shpak.

Selection against demographic stochasticity in age-structured populations. *Genetics*, 177(4):2181–2194, Dec 2007.



S. Steiner, N. Henrich, and J. Ruther.

Mating with sperm-depleted males does not increase female mating frequency in the parasitoid *Lariophagus distinguendus*. *Entomologia Experimentalis et Applicata*, 126(2):131–137, Feb. 2008.



J. E. Taylor.

The genealogical consequences of fecundity variance polymorphism. *Genetics*, 182(3):813–837, Jul 2009.



D. Waxman.

Comparison and content of the Wright-Fisher model of random genetic drift, the diffusion approximation, and an intermediate model. *J. Theoret. Biol.*, 269:79–87, 2011.