

A Wright-Fisher model with indirect selection

Ludovic Goudenège and Pierre-André Zitt

March 17, 2015

Abstract

We study a generalization of the Wright–Fisher model in which some individuals adopt a behavior that is harmful to others without any direct advantage for themselves. This model is motivated by studies of spiteful behavior in nature, including several species of parasitoid hymenoptera in which sperm-depleted males continue to mate despite not being fertile.

We first study a single reproductive season, then use it as a building block for a generalized Wright–Fisher model. In the large population limit, for male-skewed sex ratios, we rigorously derive the convergence of the renormalized process to a diffusion with a frequency-dependent selection and genetic drift. This allows a quantitative comparison of the indirect selective advantage with the direct one classically considered in the Wright–Fisher model.

From the mathematical point of view, each season is modeled by a mix between samplings with and without replacement, and analyzed by a sort of “reverse numerical analysis”, viewing a key recurrence relation as a discretization scheme for a PDE. The diffusion approximation is then obtained by classical methods.

Keywords: Wright–Fisher model; diffusion approximation; reverse numerical analysis

MSC2010: 60J20 ; 60J70; 92D15

1 Introduction: models and main results

1.1 Harmful behaviours and population genetics

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 drift¹ (its precise definition is recalled below in Section 1.4). It is simple enough that a very detailed mathematical

¹Let 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.

analysis can be performed (see for example the monographs [Dur08] or [Eth11], 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.

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 [DB06, SHR08]. 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 [Ric96, RPSWT09]. 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 (see, e.g., [Ham70], [Dio07] and [FWR01] for discussions of Hamiltonian spite).

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.

In the remainder of this introduction 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.

1.2 Basic model

In the basic model, suggested by F.-X. Dechaume-Moncharmont and M. Galipaud², 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.

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

²Personal communication.

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.

Remark 1 (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.*

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:

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

Theorem 2. *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 $w, b \geq 1$.

1.3 Large population limit

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. 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 below (see Theorem 16) 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 Ω by:

$$u(x, y, z) = \exp(-T(x, y, z)), \quad v(x, y, z) = 1 - u(x, y, z). \tag{2}$$

A heuristic derivation of the expression of u , v and T will be given below in Remark 15.

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

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

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

$$\begin{aligned} g^N : \Omega_N &\rightarrow \mathbb{R} \\ (w, b, f) &\mapsto g\left(\frac{w}{N}, \frac{b}{N}, \frac{f}{N}\right). \end{aligned} \quad (3)$$

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

$$\delta_x p(w, b, f) = p(w + 1, b, f) - p(w, b, f) \quad \delta_y p(w, b, f) = p(w, b + 1, f) - p(w, b, f). \quad (4)$$

Finally, most of the bounds we prove are uniform on specific subsets of Ω or Ω_N . For any $y_0 > 0$, and any $s < 1$, we define:

$$\begin{aligned} \Omega(y_0) &= \{(x, y, z) \in \Omega : y \geq y_0\}; \\ \Omega_N(y_0) &= \left\{ (w, b, f) \in \mathbb{N}^3 : \left(\frac{w}{N}, \frac{b}{N}, \frac{f}{N}\right) \in \Omega(y_0) \right\}; \\ \Omega(s) &= \{(x, y, z) \in \Omega : z \leq s(x + y) \text{ and } x - z \geq (1 - s)/(2 + 2s)\}; \\ \Omega_N(s) &= \left\{ (w, b, f) \in \mathbb{N}^3 : \left(\frac{w}{N}, \frac{b}{N}, \frac{f}{N}\right) \in \Omega(s) \right\}. \end{aligned}$$

Remark 3 (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 (cf. Theorem 16), while keeping an essential “stability” property (cf. the proof of the controls of errors at the end of Section 3.4).*

Now we can state the first asymptotic result.

Theorem 4. *For any $y_0 > 0$, there exists a constant $C(y_0)$ such that for all N ,*

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

where v^N is the discretization of v (see (2) and (3)). Moreover, the difference of fitness is of order $1/N$, and more precisely:

$$\forall (w, b, f) \in \Omega_N(y_0), \quad \left| p_b(w, b, f) - p_w(w, b, f) - \frac{1}{N}(\partial_x v - \partial_y v)^N(w, b, f) \right| \leq \frac{C(y_0)}{N^2}. \quad (5)$$

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 .

1.4 Multiple generations: the classical Wright–Fisher model with selection

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. See for example the monographs [Dur08, Eth11] for a much more detailed exposition; we follow here [Eth11], Section 5.2. 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.

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 $\frac{(1+\beta(N))x}{(1-x)+(1+\beta(N))x}$.

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. For an introduction to this subject, see [Eth11], [EK86] and [Ewe04].

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

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

The diffusion approximation is the following:

Theorem 5 (Wright–Fisher diffusion with selection). *In the weak selection limit, the rescaled Wright–Fisher model $(X^N)_t$ converges weakly (in the Skorokhod sense) as $N \rightarrow \infty$ to the diffusion $dX_t = \sqrt{a(x)}dB_t + b(X_t)dt$ 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 6. *If the white eggs survive better than the black ones, then $s_w(N) > s_b(N)$ so β is positive; the diffusion drifts towards $x = 1$. If black eggs are favored, β is negative and the drift is towards 0.*

1.5 A Wright–Fisher model with indirect selection

Let us now see how the basic model of Section 1.2 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 (see [CS09] for a unifying approach). Frequency-dependent coefficients may appear in such models but are often built-in in the individual-based model (see

[CS14]). Another possible extension is to make the offspring play a (game-theoretic) game after the random mating step, see e.g. [Les05]. Frequency dependence appears more naturally in [Gil74, Gil75] 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. For more details on these questions we refer to [Shp07, Tay09] and references therein. In the more complicated setting of the evolution of continuous traits, several papers [CFM06, CFM08] 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 (see [Wax11] and [MW07]) and problems arising at the boundaries.

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.

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

$$\tilde{Z}_{1,\beta}^n = \frac{(1 + \beta(N))\tilde{X}_1^n}{(1 + \beta(N))\tilde{X}_1^n + \tilde{Y}_1^n} \quad (6)$$

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 $k = 1$.

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 $(\frac{1}{n}X_k^n)_k$ with an explicit non-trivial drift towards 0. The drift and volatility are expressed in terms of the following function:

$$v_s : [0, 1] \rightarrow \mathbb{R}, \quad x \mapsto v \left(\frac{x}{1+s}, \frac{1-x}{1+s}, \frac{s}{1+s} \right), \quad (7)$$

where we recall that v is defined by (2).

Theorem 7. *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: $dX_t = \sqrt{a(x)}dB_t + b(X_t)dt$ and the corresponding generator $L = \frac{1}{2}a(x)\partial_{xx} + b(x)\partial_x$, where*

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

Remark 8. *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 9. *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 statement with explicit bounds will be given below (Lemma 24).*

A detailed study of the properties of this diffusion will be done in a forthcoming paper. Let us just stress two points as regards the comparison with the classical model of Theorem 5:

1. 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 $nv_s(x) + \mathcal{O}(1)$. Other examples of models with frequency-dependent variance effective population sizes due to polymorphism in life history traits can be found in [Gil74], [Gil75], [Shp07] or [Tay09], where the strength of selection on the alleles affecting the life history trait is also inversely proportional to the census population size.
2. 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 β : 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 β and s , there may be one or more “equilibrium” points where the drift cancels out. These cases and their interpretation in biological terms will be studied in a forthcoming paper.

Outline of the paper. In Section 2 we study the basic, single-generation model, and prove Theorem 2; we also give concentration properties for the number of reproductions. The asymptotic behaviour of p_w and p_b is studied in Section 3 where we prove Theorem 4. Finally, we prove the diffusion approximation for the multi-generation model in Section 4.

2 The single-generation model — basic properties

2.1 The advantage of being harmful

We consider here the simple model where we draw f times from an urn with w white balls and b black balls, where the white balls are removed when they are drawn and the black balls are put back in the urn.

Since similar-colored balls play the same role, p_w is the probability of a successful reproduction for a regular male, and p_b the corresponding one for a “harmful” male. Finally let $q_w(w, b, f) = 1 - p_w(w, b, f)$ and $q_b(w, b, f) = 1 - p_b(w, b, f)$ be the probabilities of not being drawn.

To prove Theorem 2, let us introduce a third function q as follows. 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:

$$q(w, b, f) = \mathbb{P}[\text{Roger is not drawn}]. \quad (8)$$

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

$$q_w(w, b, f) = q(w - 1, b, f), \quad q_b(w, b, f) = q(w, b - 1, f). \quad (9)$$

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

We use a coupling proof. Suppose that the urn 1 contains $w + b$ balls, numbered from 1 to $w + b$, where the first w balls are white, the next $b - 1$ are black and the last one is red. Urn number 2 is similar, except that the ball numbered w is black instead of white. Let (U_i) be sequence of i.i.d. random numbers, uniformly distributed on $\{1, \dots, w + b\}$. We define a joint evolution of the urns in the following way.

1. At the beginning of each step, look at the next random number; say its value is k .
2.
 - If both balls numbered k are still in their urns, choose these balls.
 - If both balls numbered k have been removed, try again with the next random number (this will only happen if the balls are both white).
 - If the ball numbered k is still in one urn but has been removed from the other, then the ball that is present is chosen. Continue looking at the next random numbers to choose a ball in the other urn.
3. At this point one ball is chosen in each urn. If any of the two is red, the process is stopped in the corresponding urn. If a chosen ball is white it is removed from its urn.
4. Repeat until the two red balls have been chosen or f draws have been made.

Each urn taken separately follows the initial process. Moreover, at any time, if the ball numbered i is still in the first urn, then it is also in the second one: indeed this is true at the beginning, and if this is true at the beginning of a step it is true at the end of the step. There are three possible situations:

- both red balls are chosen at the same time;

- the red ball is chosen in the first urn, but not in the second;
- both red balls stay untouched during the f steps.

Therefore the probability that the red ball stays untouched is smaller in the first urn than in the second urn, so

$$1 - p_b(w, b, f) = q_b(w, b, f) = q(w, b - 1, f) \leq q(w - 1, b, f) = q_w(w, b, f) = 1 - p_w(w, b, f).$$

If f is larger than 2, and w and b are larger than 1, the second case occurs with positive probability so the inequality is strict. This concludes the proof of Theorem 2.

2.2 Negative relation and concentration

In this section we prove that the total number of reproductions X and Y defined at the beginning of the Introduction are sums of “negatively related” indicators; this implies very strong concentration bounds.

In the original experiment, let us number the “white” males from 1 to w , and the “black” males from $w + 1$ to $w + b$. Let $B_i = \mathbf{1}_{\text{the } i \text{ male reproduces}}$. The total number of reproductions is given by

$$X = \sum_{i=1}^w B_i, \quad Y = \sum_{i=w+1}^b B_i. \quad (10)$$

The setting is quite close to the usual sampling from a bin with or without replacement, which leads to binomial and hypergeometric distributions. For these distributions, very strong approximation and concentration results can be proved using the fact that the indicators B_i appearing in (10) are “negatively related”: intuitively, if a certain group of males have been chosen, the others are less likely to be chosen. This approach is used in [Jan94], who refers to [BHJ92] and [JDP83] for further details on “negatively related/negatively associated” variables.

Definition 10 (Negative relation, [Jan94]). *Let I_1, \dots, I_k be indicator variables. If there exist indicator variables $J_j^{(i)}$ such that:*

- $\forall j \neq i, J_j^{(i)} \leq I_j$,
- for each i , the law of $(J_j^{(i)})_j$ is the conditional law of \mathbf{I} given $I_i = 1$,

then the variables I_i are negatively related. If they are, then $(1 - I_i)_i$ are also negatively related.

Remark 11. *For indicator variables, this corresponds to the existence of a “decreasing size-biased coupling” in the terminology of [Ros11]. However the boundedness condition used there to get concentration will not be satisfied with good constants.*

Theorem 12 (Concentration for sums of negatively related indicators). *Suppose that (I_i) are negatively related Bernoulli variables of parameter p . Let $X = \sum_{i=1}^n I_i$, and let X' be a binomial variable of parameters n and p . Then, for all $t \in \mathbb{R}$,*

$$\mathbb{E}[\exp(tX)] \leq \mathbb{E}[\exp tX'].$$

Consequently,

$$\mathbb{E} \left[|X - \mathbb{E}[X]|^3 \right] \leq 12en^{3/2} \quad (11)$$

$$\mathbb{P} [|X - \mathbb{E}[X]| \geq D] \leq \exp \left(-\frac{D^2}{4n} \right). \quad (12)$$

Proof. The key comparison of the Laplace transforms between X and the “independent version” X' comes from [Jan94, Theorem 4]. Therefore any concentration bound obtained by the usual Chernoff trick for independent variables also holds when the indicators are negatively related.

The rest of the proof is routine and is included here for completeness. Let $q = 1 - p$. For any t ,

$$\begin{aligned} \mathbb{E} [\exp(t(X - \mathbb{E}[X]))] &\leq \mathbb{E} [\exp(t(X' - \mathbb{E}[X']))] \\ &= \exp(-tnp) (pe^t + q)^n \\ &= (pe^{tq} + qe^{-tp})^n. \end{aligned}$$

Here we use a small trick borrowed from [GS01, p. 31] and bound e^x by $x + e^{x^2}$, for $x = tq$ and $x = -tp$:

$$\begin{aligned} \mathbb{E} [\exp(t(X - \mathbb{E}[X]))] &\leq (pe^{t^2q^2} + qe^{t^2p^2})^n \\ &\leq \exp(t^2n). \end{aligned}$$

The deviation inequality (12) follows by applying Markov’s exponential inequality and choosing $t = D/2n$. For the moment bound, since $\frac{|tx|^3}{3!} \leq (\exp(tx) + \exp(-tx))$,

$$\mathbb{E} [|X - \mathbb{E}[X]|^3] \leq \frac{6}{t^3} 2 \exp(t^2n)$$

Choosing $t = n^{-1/2}$ yields (11). □

Theorem 13. *Let B_i be the indicator that the i^{th} male is chosen. The indicator variables $(B_i)_{i=1, \dots, w+b}$ are negatively related.*

Proof. Define $I_i = 1 - B_i$. By the remark in Definition 10, it is enough to show that the I_i are negatively related. One may view the model as an urn occupancy problem: the $w + b$ balls become urns, in which we put f balls consecutively, not allowing more than one ball in each white urn; I_i is the event “the urn i is empty at the end”. For this type of problem, the property is standard and the $J_j^{(i)}$ may be defined explicitly in the following way. First draw the balls and record the values of the (I_i) . To define $J_j^{(i)}$, take all balls in the urn i and reassign them to the other urns, following the same procedure. Let $J_j^{(i)}$ be 1 if the urn j is empty after these reassignments. The $J_j^{(i)}$ follow the conditional distribution of $(I_1, \dots, I_{w+b}$ given $I_i = 1$, and since we only add balls to the urns j , $j \neq i$, $J_j^{(i)} \leq I_j$. □

3 The single-generation model — large population limit

3.1 Outline of the proof

The goal of this section is to prove Theorem 4 on the convergence of p_b and p_w to a continuous function v . It will be slightly easier to work on the quantity $q(w, b, f)$ defined by (8), and deduce the statements on p_w and p_b afterwards. This discrete function q approximates the function u defined by (2), and we also get convergence of the discrete differences of q to the derivatives of u :

Theorem 14. *For all $y_0 > 0$, there exists $C(y_0)$ such that, for all N , and all $(w, b, f) \in \Omega_N(y_0)$,*

$$\left| (q - u^N)(w, b, f) \right| \leq \frac{C(y_0)}{N}, \quad (13)$$

$$\left| (N\delta_x q - (\partial_x u)^N)(w, b, f) \right| \leq \frac{C(y_0)}{N}, \quad (14)$$

$$\left| (N\delta_y q - (\partial_y u)^N)(w, b, f) \right| \leq \frac{C(y_0)}{N}. \quad (15)$$

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.

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

$$q(w, b, f) = \frac{w}{w+b+1}q(w-1, b, f-1) + \frac{b}{w+b+1}q(w, b, f-1) \quad (16)$$

The main idea is then to view q as a discrete version of u , and the recurrence relation (16) as an approximation of a relation between derivatives of u . The corresponding PDE for u is derived in Section 3.2, we show in Section 3.3 that it is explicitly solvable. Knowing this, we turn to the proof of Theorem 14 in the following sections: the three convergences (13), (14) and (15) are proved respectively in Sections 3.4, 3.5 and 3.6. We show in Section 3.7 how to deduce the statements on p_w and p_b of Theorem 4 from Theorem 14. We conclude this long section by giving estimates in the same vein for second moments in Section 3.8.

3.2 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 the recurrence relation (16), we introduce $q(w, b, f)$ on the right hand side, so that discrete differences appear:

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

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

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

Now if $N = w + b + f$ goes to ∞ , 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 Ω :

$$\begin{cases} \forall (x, y, z) \in U, & u + F \cdot \nabla u = 0, \\ \forall (x, y), & u(x, y, 0) = 1, \end{cases} \quad (17)$$

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

Remark 15. *J.E. Taylor³ suggested the following heuristic justification of the expressions of T , u and v . Let X_1, \dots, X_w be the number of reproduction attempts on each of the w white balls, and Y_1, \dots, Y_b be the number of attempts on the black balls. Since there is a total of f attempts, $\sum_{i=1}^w X_i + \sum_{j=1}^b Y_j = f$ and in particular, $w\mathbb{E}[X_1] + b\mathbb{E}[Y_1] = f$. Now we make two approximations. Firstly, $\mathbb{P}[X_1 = 0] \approx \mathbb{P}[Y_1 = 0] \approx u(x, y, z)$. Secondly, the variable Y_1 should be approximately Poisson distributed: Y_1 counts successes in a large number of draws (f) that have a small chance of success. Then $\mathbb{E}[X_1] \approx 1 - u = v$, and the parameter t of the distribution of Y_1 satisfies $e^{-t} = u$, so $t = -\log(1 - v)$. Inserting this in the equation on expectations yields $xv - y \log(1 - v) = z$, which is another form of the equations (1) and (2) defining v .*

A complete justification of these arguments, and in particular of the Poisson approximation, should be possible but could be quite involved, since the dependence between the draws is not easy to take into account.

3.3 Resolution of the PDE

This first order PDE (17) 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:

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

³Personal communication.

Therefore $h(t) = h(0) \exp(t)$. Suppose $T = T(x_0, y_0, z_0)$ is a solution of (1), i.e. $z(T) = 0$. Then $h(T) = u(M(T)) = 1$ thanks to the boundary condition. Finally :

$$u(x_0, y_0, z_0) = h(0) = u(M(T)) \exp(-T) = \exp(-T(x_0, y_0, z_0)). \quad (18)$$

Theorem 16 (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 defined by (18) 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) ,*

$$|u(x, y, z)| \leq C(y_0), \quad |\partial_i u(x, y, z)| \leq C(y_0), \quad |\partial_i \partial_j u(x, y, z)| \leq C(y_0). \quad (19)$$

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

Proof. If y is strictly positive, $\phi : t \mapsto x(e^{-t} - 1) - yt + z$ is a strictly decreasing smooth function such that $\phi(0) = z$ and $\phi(z/y) < 0$. Therefore T is unique and depends smoothly on x, y, z by the implicit function theorem. Its derivatives are given by:

$$\partial_x T = \frac{e^{-T} - 1}{xe^{-T} + y}; \quad \partial_y T = \frac{-T}{xe^{-T} + y}; \quad \partial_z T = \frac{1}{xe^{-T} + y}.$$

On $\Omega(y_0)$, T is positive and smaller than $1/y_0$, therefore these quantities are bounded. The same is true for the higher order derivatives.

If $s < 1$, recall that on $\Omega(s)$,

$$z \leq s(x + y) \quad x - z \geq (1 - s)/(2 + 2s). \quad (20)$$

By the first condition, we obtain $\phi(\ln(1/(1 - s))) \leq y(\ln(1 - s) + s) \leq 0$ which implies that $T \leq \ln(1/(1 - s))$. Together with the second condition, this implies that the denominator $xe^{-T} + y \geq xe^{-T} \geq x(1 - s) \geq (1 - s)^2/(2 + 2s)$. This proves the claimed bounds. \square

3.4 Convergence

In this section we prove (13). Let u be the solution (18) of the continuous PDE, and u^N its discretization defined by (3). If the recurrence relation (16) can be seen as a numerical scheme for the resolution of the PDE (17), u^N should approximately satisfy (16). Define R_N to be the corresponding difference:

$$R_N(w, b, f) = u^N(w, b, f) - \frac{w}{w + b + 1} u^N(w - 1, b, f - 1) - \frac{b}{w + b + 1} u^N(w, b, f - 1). \quad (21)$$

Proposition 17. *For all $y_0 > 0$, there exists $C(y_0)$ such that for all N ,*

$$\forall (w, b, f) \in \Omega_N(y_0), \quad |R_N(w, b, f)| \leq \frac{C(y_0)}{N^2}.$$

If $s < 1$, a similar bound holds uniformly on $\Omega_N(s)$.

Proof. Let $m_N(w, b, f)$ be the sup of the second derivatives of u on the cell $[(w \pm 1)/N] \times [(b \pm 1)/N] \times [(f \pm 1)/N]$. Let $\mathbf{x}_N = (w/N, b/N, f/N)$, so that $u^N(w, b, f) = u(\mathbf{x}_N)$. Multiply (21) by $(w + b + 1)$ and apply Taylor's formula:

$$\begin{aligned} (w + b + 1)R_N(w, b, f) &= (w + b + 1)u(\mathbf{x}_N) - w \left(u(\mathbf{x}_N) - \frac{1}{N} \partial_x u(\mathbf{x}_N) - \frac{1}{N} \partial_z u(\mathbf{x}_N) \right) \\ &\quad - b \left(u(\mathbf{x}_N) - \frac{1}{N} \partial_z u(\mathbf{x}_N) \right) + (w + b + 1)\epsilon(w, b, f) \end{aligned}$$

where $|\epsilon(w, b, f)| \leq \frac{1}{N^2} m_N(w, b, f)$. So

$$(w + b + 1)R_N(w, b, f) = u(\mathbf{x}_N) + w \left(\frac{1}{N} \partial_x u(\mathbf{x}_N) + \frac{1}{N} \partial_z u(\mathbf{x}_N) \right) + b \frac{1}{N} \partial_z u(\mathbf{x}_N) + (w + b + 1)\epsilon(w, b, f).$$

Since u solves the PDE, all terms vanish except the last one, so

$$|R_N(w, b, f)| \leq \frac{m_N(w, b, f)}{N^2}.$$

The controls on the derivatives of u given by Theorem 16 show that m_N is bounded by some $C(y_0)$ on $\Omega_N(y_0)$, and by some $C(s)$ on $\Omega_N(s)$: this concludes the proof. \square

Now let $e_N(w, b, f)$ be the difference $q(w, b, f) - u^N(w, b, f)$. By the recurrence relation (16) and the definition (21) of R_N , for $w \geq 1$ and $f \geq 1$ we get:

$$e_N(w, b, f) = \frac{w}{w + b + 1} e_N(w - 1, b, f - 1) + \frac{b}{w + b + 1} e_N(w, b, f - 1) - R_N(w, b, f).$$

This still holds for $w = 0$ if we define $e_N(-1, b, f) = 0$.

Now define $\overline{e}_N(f) = \max\{|e_N(w, b, f)| : (w, b) \in \mathbb{N}^2, (w, b, f) \in \Omega_N(y_0)\}$. The key fact is that, if (w, b, f) is in $\Omega_N(y_0)$, the same is true for $(w - 1, b, f - 1)$ and $(w, b, f - 1)$. Therefore:

$$\begin{aligned} \overline{e}_N(f) &\leq \overline{e}_N(f - 1) + \max\{R_N(w, b, f) : w, b; (w, b, f) \in \Omega_N(y_0)\} \\ &\leq \overline{e}_N(f - 1) + \frac{C(y_0)}{N^2}. \end{aligned}$$

By induction, since $f \leq N$,

$$\overline{e}_N(f) \leq \overline{e}_N(0) + \frac{C(y_0)}{N}.$$

Since $e_N(w, b, 0) = 0$, we are done.

To prove the bounds on $\Omega(s)$, the strategy is exactly the same. Once more, the crucial step is to remark that $(w - 1, b, f - 1)$ and $(w, b, f - 1)$ belong to $\Omega_N(s)$ whenever $(w, b, f) \in \Omega_N(s)$: this stability is the reason behind the very definition of $\Omega(s)$.

3.5 Derivative in the x direction

Let us now prove the convergence of the (renormalized) finite differences of q to the derivatives of u . We proceed in three steps:

1. find a recurrence relation for the finite differences;
2. find a PDE for the derivative;
3. use the PDE to show that the discretization of the derivatives almost follows the same recurrence relation as the finite differences.

We begin by the convergence of the derivatives in the x direction. Define $u_x = \partial_x u$, and recall that $\delta_x q(w, b, f)$ is the finite difference:

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

Step 1. In order to obtain a recurrence relation for $\delta_x q$, starting from its definition, we apply (16) two times to $q(w+1, b, f)$ and $q(w, b, f)$:

$$\begin{aligned}
\delta_x q(w, b, f) &= \frac{w+1}{w+b+2}q(w, b, f-1) + \frac{b}{w+b+2}q(w+1, b, f-1) \\
&\quad - \frac{w}{w+b+1}q(w-1, b, f-1) - \frac{b}{w+b+1}q(w, b, f-1) \\
&= \frac{w}{w+b+1}\delta_x q(w-1, b, f-1) + \frac{b}{w+b+2}\delta_x q(w, b, f-1) \\
&\quad + \left(\frac{w+1}{w+b+2} - \frac{w}{w+b+1} \right) q(w, b, f-1) \\
&\quad + \left(\frac{b}{w+b+2} - \frac{b}{w+b+1} \right) q(w, b, f-1) \\
&= \frac{w}{w+b+1}\delta_x q(w-1, b, f-1) + \frac{b}{w+b+2}\delta_x q(w, b, f-1) \\
&\quad + \frac{1}{(w+b+1)(w+b+2)}q(w, b, f-1). \tag{22}
\end{aligned}$$

Step 2. Now let us find a PDE for u_x . Recall that $u_x = \partial_x u$. Since $u + x\partial_x u + (x+y)\partial_z u = 0$, u_x satisfies:

$$2u_x + x\partial_x(u_x) + \partial_z u + (x+y)\partial_z u_x = 0.$$

Plugging the first equation into the second gives:

$$2u_x + x\partial_x u_x - \frac{1}{x+y}u - \frac{x}{x+y}u_x + (x+y)\partial_z u_x = 0,$$

which simplifies to:

$$\frac{x+2y}{x+y}u_x + x\partial_x u_x + (x+y)\partial_z u_x = \frac{1}{x+y}u. \tag{23}$$

Step 3. Let u_x^N be the discretization of u_x . This function should approximately satisfy the same relation as $N\delta_x q$, i.e. the product of Equation (22) by N . Denote by R_N the error in this approximation, i.e. R_N is such that:

$$\begin{aligned}
u_x^N(w, b, f) &= \frac{w}{w+b+1}u_x^N(w-1, b, f-1) + \frac{b}{w+b+2}u_x^N(w, b, f-1) \\
&\quad + \frac{N}{(w+b+1)(w+b+2)}q(w, b, f-1) + R_N(w, b, f). \tag{24}
\end{aligned}$$

The error $e_N = u_x^N - N\delta_x q$ satisfies:

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

so the same proof as before applies, provided we show that

- R_N is $\mathcal{O}(N^{-2})$;
- $e_N(w, b, 0)$ is small.

Lemma 18 (R_N is small). *For any y_0 , there exists a $C(y_0)$ such that*

$$\forall (w, b, f) \in \Omega_N(y_0), \quad R_N(w, b, f) \leq \frac{C(y_0)}{N^2}. \quad (25)$$

The same holds uniformly on $\Omega_N(s)$ if $s < 1$.

Proof. Multiply (24) by $(w + b + 1)$ and use Taylor’s formula:

$$\begin{aligned} & (w + b + 1)u_x^N(w, b, f) \\ &= w \left(u_x^N(w, b, f) - \frac{1}{N}(\partial_x u_x)^N(w, b, f) - \frac{1}{N}(\partial_z u_x)^N(w, b, f) \right) \\ & \quad + b \left(1 - \frac{1}{w + b + 2} \right) \left(u_x^N(w, b, f) - \frac{1}{N}(\partial_z u_x)^N(w, b, f) \right) \\ & \quad + \frac{N}{w + b + 2}q(w, b, f - 1) + (w + b + 1)(R_N(w, b, f) + \epsilon(w, b, f)), \end{aligned}$$

where $|\epsilon(w, b, f)| \leq m_N(w, b, f)$. Gather all the u_x^N terms on the left hand side.

$$\begin{aligned} \frac{w + 2b + 2}{w + b + 2}u_x^N(w, b, f) &= -w \left(\frac{1}{N}(\partial_x u_x)^N(w, b, f) + \frac{1}{N}(\partial_z u_x)^N(w, b, f) \right) \\ & \quad - \frac{b}{N} \left(1 - \frac{1}{w + b + 2} \right) \partial_z u_x^N(w, b, f) \\ & \quad + \frac{N}{w + b + 2}q(w, b, f - 1) + (w + b + 1)(R_N(w, b, f) + \epsilon(w, b, f)). \end{aligned}$$

To use the fact that u_x satisfies (23) we isolate the relevant terms:

$$\begin{aligned} & \frac{w + 2b}{w + b}u_x^N(w, b, f) - \frac{2b}{(w + b)(w + b + 2)}u_x^N(w, b, f) \\ &= -w \left(\frac{1}{N}(\partial_x u_x)^N(w, b, f) + \frac{1}{N}(\partial_z u_x)^N(w, b, f) \right) \\ & \quad - \frac{b}{N}(\partial_z u_x)^N(w, b, f) + \frac{b}{N} \frac{1}{w + b + 2}(\partial_z u_x)^N(w, b, f) \\ & \quad + \frac{N}{w + b}u^N(w, b, f) + \frac{N}{w + b}(q(w, b, f - 1) - u^N(w, b, f)) \\ & \quad - \frac{2Nq(w, b, f - 1)}{(w + b)(w + b + 2)} + (w + b + 1)(R_N(w, b, f) + \epsilon(w, b, f)). \end{aligned}$$

Thanks to (23) applied at the point $(w/N, b/N, f/N)$, we obtain:

$$\begin{aligned} & - \frac{2b}{(w + b)(w + b + 2)}u_x^N(w, b, f) \\ &= \frac{b}{N} \frac{1}{w + b + 2}(\partial_z u_x)^N(w, b, f) \\ & \quad + \frac{N}{w + b}(q(w, b, f - 1) - u^N(w, b, f)) - \frac{2Nq(w, b, f - 1)}{(w + b)(w + b + 2)} \\ & \quad + (w + b + 1)(R_N(w, b, f) + \epsilon(w, b, f)). \end{aligned}$$

Isolating R_N in this equation and using the fact that $b \geq Ny_0$, the bounds $|u_x^N| \leq 1$, $|\partial_z u_x^N| \leq C(y_0)$ as stated in Theorem 16, and the approximation result on q (Equation (13)), we get the bound (25). On $\Omega(s)$ the proof is the same, replacing the lower

bound on b on the denominator by the control

$$w = Nx \geq N \frac{1-s}{2+2s}.$$

This concludes the proof of the lemma. \square

To conclude the proof of (14), we need only consider the base case $f = 0$. Since $q(\cdot, \cdot, 0)$ is identically 1, $\delta_x q$ is zero for $f = 0$. Similarly u is identically 1 so its x -derivative is 0, so $e_N(w, b, 0) = 0$, and (14) follows by the same induction as before.

3.6 The other derivatives

Let us now turn to the convergence of the y -derivative u_y . This function satisfies the PDE:

$$u_y + x\partial_x u_y + (x+y)\partial_z u_y = -\frac{1}{x+y}u - \frac{x}{x+y}u_x. \quad (26)$$

Note that the right hand side depends on u and u_x , for which we have already proved approximation results.

Recall that $\delta_y q(w, b, f) = q(w, b+1, f) - q(w, b, f)$. Using the recurrence relation (16) for q we find first that

$$\begin{aligned} & \delta_y q(w, b, f) \\ &= \frac{w}{w+b+2}q(w-1, b+1, f-1) + \frac{b+1}{w+b+2}q(w, b+1, f-1) \\ & \quad - \frac{w}{w+b+1}q(w-1, b, f-1) - \frac{b}{w+b+1}q(w, b+1, f-1) \\ &= \frac{w}{w+b+1}\delta_y q(w-1, b, f-1) + \frac{b}{w+b+1}\delta_y q(w, b, f-1) \\ & \quad - \frac{w}{(w+b+1)(w+b+2)}q(w-1, b+1, f-1) \\ & \quad + \frac{w+1}{(w+b+1)(w+b+2)}q(w, b+1, f-1) \\ &= \frac{w}{w+b+1}\delta_y q(w-1, b, f-1) + \frac{b}{w+b+1}\delta_y q(w, b, f-1) \\ & \quad + \frac{w}{(w+b+1)(w+b+2)}\delta_x q(w-1, b+1, f-1) \\ & \quad + \frac{1}{(w+b+1)(w+b+2)}q(w, b+1, f-1). \end{aligned}$$

Once more, the discretization u_y^N of u_y should behave approximately like $N\delta_y q$. Define R_N to be the error in this approximation, i.e. R_N is such that:

$$\begin{aligned} u_y^N(w, b, f) &= \frac{w}{w+b+1}u_y^N(w-1, b, f-1) + \frac{b}{w+b+1}u_y^N(w, b, f-1) \\ & \quad + \frac{wN}{(w+b+1)(w+b+2)}\delta_x q(w-1, b+1, f-1) \\ & \quad + \frac{N}{(w+b+1)(w+b+2)}q(w, b+1, f-1) \\ & \quad + R_N(w, b, f). \end{aligned}$$

To study R_N , multiply both sides by $(w + b + 1)$, and use Taylor’s formula:

$$\begin{aligned} & (w + b + 1)u_y^N(w, b, f) \\ &= wu_y^N(w, b, f) + bu_y^N(w, b, f) - \frac{w}{N}((\partial_x + \partial_z)u_y)^N(w, b, f) - \frac{b}{N}(\partial_z u_y)^N(w, b, f) \\ & \quad + \frac{wN}{w + b + 2}\delta_x q(w - 1, b, f - 1) + \frac{N}{w + b + 2}q(w, b + 1, f - 1) \\ & \quad + (w + b + 1)R_N(w, b, f) + \epsilon(w, b, f), \end{aligned}$$

where $\epsilon = \mathcal{O}(N^{-1})$ (uniformly on $\Omega(y_0)$ and on $\Omega(s)$). The term $(w + b)(u_y)^N$ cancels out. The remaining terms almost cancel out thanks to (26), and we are left with

$$\begin{aligned} (w + b + 1)R_N(w, b, f) &= \mathcal{O}(1/N) + \left(\frac{N}{w + b + 2}q(w, b + 1, f - 1) - \frac{N}{w + b}u^N(w, b, f) \right) \\ & \quad + \left(\frac{wN}{w + b + 2}\delta_x q(w - 1, b + 1, f - 1) - \frac{w}{w + b}u_x^N(w, b, f) \right). \end{aligned}$$

Using the approximation results (13), (14), and the fact that $b \geq Ny_0$ on $\Omega_N(y_0)$, or that $w \geq N\frac{1-s}{2+2s}$ on $\Omega_N(s)$, we can prove that

$$|R_N(w, b, f)| \leq C(y_0)N^{-2}.$$

The last step is the same as before: the difference $e_N = N\delta_y q - u_y^N$ satisfies the nice recurrence relation

$$e_N(w, b, f) = \frac{w}{w + b + 1}e_N(w - 1, b, f - 1) + \frac{b}{w + b + 1}e_N(w, b, f - 1) - R_N(w, b, f).$$

For the base case ($f = 0$), e_N is identically zero, and we get by induction:

$$\max \{|e_N(w, b, f)| : (w, b) \text{ such that } (w, b, f) \in \Omega_N(y_0)\} \leq \frac{C(y_0)f}{N^2},$$

which proves (15) since $f \leq N$. The proof is similar on $\Omega_N(s)$.

3.7 Proof of Theorem 4

Let us see how the statements of Theorem 4 regarding p_b and p_w may be deduced from Theorem 14. The three proofs being similar, we only consider the last equation (5), that is, we prove

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

Recall that p_w and p_b can be expressed in terms of q : by (9),

$$p_w(w, b, f) = 1 - q_w(w, b, f) = 1 - q(w - 1, b, f), \quad (27)$$

$$p_b(w, b, f) = 1 - q_b(w, b, f) = 1 - q(w, b - 1, f). \quad (28)$$

First write everything in terms of q and u , recalling that $u = 1 - v$:

$$\begin{aligned} & p_b(w, b, f) - p_w(w, b, f) - \frac{1}{N}(\partial_x v - \partial_y v)^N(w, b, f) \\ &= q(w - 1, b, f) - q(w, b - 1, f) + \frac{1}{N}(\partial_x u - \partial_y u)^N(w, b, f) \\ &= q(w - 1, b, f) - q(w, b, f) + \frac{1}{N}(\partial_x u)^N(w, b, f) \end{aligned} \quad (29)$$

$$+ q(w, b, f) - q(w, b - 1, f) - \frac{1}{N}(\partial_y u)^N(w, b, f). \quad (30)$$

The absolute value of the last line (30) satisfies:

$$\begin{aligned} & \left| q(w, b, f) - q(w, b - 1, f) - \frac{1}{N} (\partial_y u)^N(w, b, f) \right| \\ & \leq \frac{1}{N} \left| N \delta_y q(w, b - 1, f) - (\partial_y u)^N(w, b - 1, f) \right| \\ & \quad + \frac{1}{N} \left| (\partial_y u)^N(w, b - 1, f) - (\partial_y u)^N(w, b, f) \right|. \end{aligned}$$

Fix $0 < y'_0 < y_0$ to ensure that $(w, b - 1, f)$ is in $\Omega_N(y'_0)$ when (w, b, f) is in $\Omega_N(y'_0)$. By (15) the first term is bounded by $\frac{C(y'_0)}{N^2}$. The controls on u from Theorem 16 imply that the second term is also bounded by $C(y'_0)/N^2$. The same arguments may be applied to the terms in (29). This concludes the proof of (5).

3.8 Second moments

In order to derive the diffusion limit, we will need information on the covariance structure of the couple (X, Y) . This information will be deduced in Section 4.3 from estimates on the following variant of the function q .

Definition 19. *For any positive integers w, b and f , we denote by $\tilde{q}(w, b, f)$ the probability that in an urn composed of w white balls, b black balls and 2 red balls, the 2 red balls are not drawn after f trials.*

By conditioning, we see that \tilde{q} satisfies the recurrence relation:

$$\tilde{q}(w, b, f) = \frac{w}{w + b + 2} \tilde{q}(w - 1, b, f - 1) + \frac{b}{w + b + 2} \tilde{q}(w, b - 1, f - 1).$$

As before we prove that \tilde{q} converges in some sense to a limit function \tilde{u} . The same heuristic reasoning as before leads to the candidate PDE:

$$-\frac{2}{x + y} \tilde{u} - \frac{x}{x + y} (\partial_x \tilde{u} + \partial_z \tilde{u}) - \frac{y}{x + y} \partial_z \tilde{u} = 0$$

which we rewrite as

$$2\tilde{u} + x\partial_x \tilde{u} + (x + y)\partial_z \tilde{u} = 0,$$

with the boundary condition

$$\tilde{u}(x, y, 0) = 1.$$

The only difference between this equation and the PDE (17) is that F is replaced by $F/2$. We solve this new equation in the same way. The characteristics are:

$$\begin{cases} x(t) = x_0 e^{-t/2} \\ y(t) = y_0 \\ z(t) = x_0(e^{-t/2} - 1) - y_0 t/2 + z_0. \end{cases}$$

The \tilde{T} that satisfies $z(\tilde{T}) = 0$ is just $\tilde{T} = 2T$, so the solution \tilde{u} is given by:

$$\tilde{u}(x_0, y_0, z_0) = \exp(2T(x_0, y_0, z_0)) = u^2(x_0, y_0, z_0).$$

Following the same strategy as before, we prove:

Proposition 20 (Asymptotics of \tilde{q}). *For any y_0 there exists $C(y_0)$ such that for all N and for all $(w, b, f) \in \Omega_N(y_0)$,*

$$\begin{aligned} \left| (\tilde{q} - (u^N)^2)(w, b, f) \right| &\leq \frac{C(y_0)}{N}, \\ \left| (\tilde{q}(w, b - 1, f) - \tilde{q}(w - 1, b, f)) - \frac{2}{N}(u(\partial_x - \partial_y)u)^N(w, b, f) \right| &\leq \frac{C(y_0)}{N}. \end{aligned}$$

Similar bounds hold uniformly on $\Omega_N(s)$ if $s < 1$.

4 The multi-generation model

4.1 Main line of the proof

To prove the diffusion limit stated in Theorem 7, we follow the presentation of Durrett in [Dur96]. 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:

$$\begin{aligned} a^n(x) &= n \mathbf{Var}_x(X_1^n), & (31) \\ b^n(x) &= n (\mathbb{E}_x[X_1^n] - x), & (32) \end{aligned}$$

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.

Theorem 21 (Diffusion limit, [Dur96] Theorem 8.7.1 and Lemma 8.8.2). *Suppose that the following three conditions hold.*

1. *The infinitesimal mean and variance converge uniformly:*

$$\begin{aligned} \limsup_n \sup_{x \in \mathcal{S}_n} |a^n(x) - a(x)| &= 0, \\ \limsup_n \sup_{x \in \mathcal{S}_n} |b^n(x) - b(x)| &= 0. \end{aligned}$$

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 22. *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, Theorem 7 will follow once we prove that the martingale problem is well posed and we show the following estimates.

Proposition 23 (Infinitesimal mean and variance). *The following estimates hold:*

$$a_n(x) = \frac{x(1-x)}{v_s(x)} + \mathcal{O}(1/\sqrt{n}), \quad (33)$$

$$b_n(x) = x(1-x) \left(\beta - \frac{v'_s(x)}{v_s^2(x)} \right) + \mathcal{O}(1/\sqrt{n}), \quad (34)$$

$$c_n(x) = \mathcal{O}(1/\sqrt{n}), \quad (35)$$

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 $s < 1$.

Outline of the section. The remainder of this section is organized as follows. We verify in Section 4.2 that the martingale problem is well posed. In Section 4.3 we use the convergence results of the single generation model to study the “first step” and get information on the asymptotics of the random number of reproductions. Focusing first on the case where $\beta = 0$ (i.e. there is no “direct” fitness advantage), we prove the formula (34) for the infinitesimal mean in Section 4.4, postponing an estimate of a remainder term to Section 4.5. The infinitesimal variance formula (33) and the control (35) on the higher moments of the jumps are proved in sections 4.6 and 4.7, still in the case $\beta = 0$. Finally we show in Section 4.8 how to recover all these results in the case where β is arbitrary.

4.2 The martingale problem is well posed

Recall the definition (7) of the function v_s :

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

where $T \left(\frac{x}{1+s}, \frac{1-x}{1+s}, \frac{s}{1+s} \right)$ satisfies

$$x \left(1 - \exp \left(-T \left(\frac{x}{1+s}, \frac{1-x}{1+s}, \frac{s}{1+s} \right) \right) \right) + (1-x) T \left(\frac{x}{1+s}, \frac{1-x}{1+s}, \frac{s}{1+s} \right) = s \quad (36)$$

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 24 (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 $x \in [0, 1]$, $s(1 - s)^2(e^{-s} + s - 1) \leq v''_s(x) \leq \frac{2se^{-s}(-s - \log(1-s))}{(1-s)^3}$.

Proof. From now and only in this proof, we write v (resp. T) instead of $v_s(x)$ (resp. $T(x/1 + s, 1 - x/1 + s, s/1 + s)$) for convenience. Since $1 - e^{-t} \leq t$ for all $t \in \mathbb{R}_+$,

$$x(1 - e^{-T}) + (1 - x)T \leq T$$

which implies $s \leq T$ by the definition (36) of T . Then we have

$$xs + (1 - x)T \geq xs + (1 - x)s = s = xv + (1 - x)T,$$

which implies $v \leq s$ for $x > 0$. Since $T(0, 1/1 + s, s/1 + s) = s$, we have $v_s(0) = 1 - e^{-s} \leq s$. Moreover $s \leq T$ obviously implies $1 - e^{-s} \leq 1 - e^{-T} = v$, proving the first point. Finally if $x < 1$ then $T > 0$ thus $s < T$ and $v_s(x) < s$ for all $x \in [0, 1)$.

Rewriting (36) in terms of v yields the relation

$$xv - (1 - x)\log(1 - v) = s. \quad (37)$$

Differentiating this formula and isolating v' one gets $v' = \frac{(-v - \log(1-v))(1-v)}{1-xv}$; using (37) to get rid of the logarithm yields

$$v' = \frac{(1 - v)}{1 - xv} \cdot \frac{s - v}{1 - x}, \quad (38)$$

proving $v'_s(x) > 0$ for all $x \in [0, 1)$.

Since $t \mapsto -t - \log(1 - t)$ is nondecreasing on $[0, 1]$, we obtain for $0 < s < 1$

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

which gives the following bounds (uniformly with respect to x):

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

Let us now turn to the second derivative. Take the logarithm of (38) and differentiate:

$$[\log(v')] = \frac{v''}{v'} = \frac{-v'}{1 - v} + \frac{-v'}{s - v} + \frac{v}{1 - xv} + \frac{xv'}{1 - xv} + \frac{1}{1 - x}.$$

Using the expression (38) for v' , it is easy to see that the sum of the first and fourth terms is $-(s - v)/(1 - xv)^2$ and the sum of the second and fifth terms is $v/(1 - xv)$. Therefore the whole sum is simply

$$\frac{v''}{v'} = -\frac{s - v}{(1 - xv)^2} + \frac{v}{1 - xv} + \frac{v}{1 - xv} = \frac{-2xv^2 + 3v - s}{(1 - xv)^2}.$$

From this expression and the bounds on v' , the upper bound on v'' is easily obtained. Moreover, since

$$3v - 2xv^2 - s - s(1 - s) \geq -2(v - 3/4)^2 + (s - 1)^2 + 1/8$$

the lower bound $s(1-s)$ on v''/v' will follow if we show that $-2(v-3/4)^2 + (s-1)^2 + 1/8$ is positive. This quantity is minimal if $|v-3/4|$ is maximal. Since v is nondecreasing, the maximal value of $|v-3/4|$ is attained at $x=0$ or $x=1$, for which $v=1-e^{-s}$ or $v=s$. When $v=1-e^{-s}$, we have $3v-2v^2-s-s(1-s)=3-3e^{-s}-2(1-e^{-s})^2-s-s(1-s)$ which is positive for all $s<1$. When $v=s$, we have $3v-2v^2-s-s(1-s)=s(1-s)\geq 0$. This concludes the proof of the lower bound for v'' . \square

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 [Dur96, Theorem 5.3.3].

Theorem 25 (Yamada-Watanabe). *Let $dX_t = \sqrt{a(x)}dB_t + b(X_t)dt$ be a SDE such that*

(i) *there exists a positive increasing function ρ on $(0, +\infty)$ such that*

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

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

(ii) *there exists a positive increasing concave function κ on $(0, +\infty)$ such that*

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

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

Then pathwise uniqueness holds for the SDE.

For \sqrt{a} , thanks to the previous bounds in Lemma 24 and the elementary inequality $|\sqrt{c} - \sqrt{d}| \leq \sqrt{|c-d|}$, valid for all $(c, d) \in \mathbb{R}_+^2$, there exists a constant C depending only on s such that

$$\begin{aligned} \left| \sqrt{a(x)} - \sqrt{a(y)} \right| &\leq \left| \sqrt{\frac{x(1-x)}{v_s(x)}} - \sqrt{\frac{y(1-y)}{v_s(x)}} \right| + \left| \sqrt{\frac{y(1-y)}{v_s(x)}} - \sqrt{\frac{y(1-y)}{v_s(y)}} \right| \\ &\leq \frac{|\sqrt{x(1-x)} - \sqrt{y(1-y)}|}{\sqrt{v_s(x)}} + \sqrt{\frac{y(1-y)}{v_s(x)v_s(y)}} \left| \sqrt{v_s(y)} - \sqrt{v_s(x)} \right| \\ &\leq \frac{|\sqrt{x(1-x)} - \sqrt{y(1-y)}|}{\sqrt{v_s(x)}} + \sqrt{\frac{y(1-y)}{v_s(x)v_s(y)}} \sup_{z \in [0,1]} \frac{|v'_s(z)|}{2\sqrt{v_s(z)}} |x-y| \\ &\leq C\sqrt{|x-y|}\sqrt{|1-x-y|} + C|x-y| \leq 2C\sqrt{|x-y|}. \end{aligned}$$

Thus the first item holds with $\rho(u) = 2C\sqrt{u}$. For the drift b , we have

$$\begin{aligned} |b(x) - b(y)| &\leq \beta|x(1-x) - y(1-y)| + \left| x(1-x)\frac{v'_s(x)}{v_s^2(x)} - y(1-y)\frac{v'_s(y)}{v_s^2(y)} \right| \\ &\leq \beta|x-y||1-x-y| + \frac{v'_s(x)}{v_s^2(x)}|x(1-x) - y(1-y)| + y(1-y)\left| \frac{v'_s(x)}{v_s^2(x)} - \frac{v'_s(y)}{v_s^2(y)} \right| \\ &\leq (\beta + C)|x-y| + y(1-y) \sup_{z \in [0,1]} \frac{|v''_s(z)v_s^2(z) - 2v'_s(z)v_s(z)v'_s(z)|}{v_s^4(z)} |x-y| \\ &\leq (\beta + 2C)|x-y|, \end{aligned}$$

which proves the second item by setting $\kappa(u) = (\beta + 2C)u$.

4.3 The number of reproductions

In this section we use the results from the previous section to study the “first step” of each generation, getting information on the asymptotics of the random number of reproductions.

Notation. From now on, 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.

Proposition 26 (Moments of (\tilde{X}, \tilde{Y})). *The moments of (\tilde{X}, \tilde{Y}) have the following asymptotic behaviour:*

$$\begin{aligned} \tilde{x} &= \mathbb{E}_x [\tilde{X}] = nxv_s(x) + \mathcal{O}(1), & \tilde{y} &= \mathbb{E}_x [\tilde{Y}] = n(1-x)v_s(x) + \mathcal{O}(1), \\ \mathbf{Var}_x (\tilde{X}) &= \mathcal{O}(n), & \mathbf{Var}_x (\tilde{Y}) &= \mathcal{O}(n), \\ \mathbf{Cov}_x (\tilde{X}, \tilde{Y}) &= \mathcal{O}(n), \end{aligned}$$

Moreover,

$$\begin{aligned} -b\mathbf{Var}_x (\tilde{X}) + (w-b)\mathbf{Cov}_x (\tilde{X}, \tilde{Y}) + w\mathbf{Var}_x (\tilde{Y}) &= wb(v'_s(x)(v_s(x)-1) + \mathcal{O}(1/n)) \\ b^2\mathbf{Var}_x (\tilde{X}) - 2wb\mathbf{Cov}_x (\tilde{X}, \tilde{Y}) + w^2\mathbf{Var}_x (\tilde{Y}) &= wbn(v_s(x)(1-v_s(x)) + \mathcal{O}(1/n)). \end{aligned}$$

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 $s < 1$). Finally,

$$\begin{aligned} \mathbb{E}_x \left[\left| \tilde{X} - \mathbb{E} [\tilde{X}] \right|^3 \right] &= \mathcal{O}(n^{3/2}) \\ \mathbb{E}_x \left[\left| \tilde{Y} - \mathbb{E} [\tilde{Y}] \right|^3 \right] &= \mathcal{O}(n^{3/2}), \end{aligned}$$

where the \mathcal{O} are uniform on $x \in [0, 1]$.

Remark 27. Getting the exact value of the leading term for the second moments does not seem easy; we will only need a control on the particular linear combinations that appear in the second block of equations.

We begin with a lemma. Define p_{ww} to be the probability that two given different white balls are drawn, and define p_{wb} , p_{bb} similarly.

Lemma 28. For all y_0 , there exists a $C(y_0)$ such that, if $(w, b, f) \in \Omega_N(y_0)$,

$$\left| (p_{ww} - v^2)(w, b, f) \right| \leq \frac{C(y_0)}{N}, \quad \left| (p_{wb} - v^2)(w, b, f) \right| \leq \frac{C(y_0)}{N}, \quad \left| (p_{bb} - v^2)(w, b, f) \right| \leq \frac{C(y_0)}{N}.$$

Moreover, the differences are of order $1/N$ and are given by:

$$\begin{aligned} \left| \left(p_{bb} - p_{wb} - \frac{1}{N} ((2v-1)(\partial_x - \partial_y)v)^N \right) (w, b, f) \right| &\leq \frac{C(y_0)}{N^2}, \\ \left| \left(p_{wb} - p_{ww} - \frac{1}{N} ((2v-1)(\partial_x - \partial_y)v)^N \right) (w, b, f) \right| &\leq \frac{C(y_0)}{N^2}. \end{aligned}$$

The same bounds hold uniformly on $\Omega_N(s)$ if $s < 1$.

Proof. To compute these quantities, let $q_{ww} = \mathbb{P}[\text{neither } B_1 \text{ nor } B_2 \text{ are drawn}]$, and define q_{wb}, q_{bb} similarly. As before, the probability q_{ww} does not depend on the color of the two balls, but only on the composition of the remainder of the urn. In terms of the quantity \tilde{q} introduced in Definition 19, we have:

$$\begin{aligned} q_{ww}(w, b, f) &= \tilde{q}(w-2, b, f) & q_{wb}(w, b, f) &= \tilde{q}(w-1, b-1, f), \\ q_{bb}(w, b, f) &= \tilde{q}(w, b-2, f). \end{aligned}$$

Going back to the probabilities of reproduction is easy. Since for any events A and B , $\mathbb{P}[A^c \cap B^c] + \mathbb{P}[A] + \mathbb{P}[B] = 1 + \mathbb{P}[A \cap B]$, we get

$$\begin{aligned} p_{ww} &= q_{ww} - 1 + 2p_w \\ p_{wb} &= q_{wb} - 1 + p_w + p_b \\ p_{bb} &= q_{bb} - 1 + 2p_b. \end{aligned}$$

The result follows using the previous approximations on p_w and p_b from Theorem 4 and the results on \tilde{q} (Proposition 20). \square

Proof of Proposition 26. The variance and covariance of \tilde{X} and \tilde{Y} are easily computed in terms of these quantities.

$$\begin{aligned} \mathbf{Var}_x(\tilde{X}) &= \mathbf{Var}\left(\sum_{i=1}^w \mathbf{1}_{B_i}\right) = w\mathbf{Var}_x(\mathbf{1}_{B_1}) + w(w-1)\mathbf{Cov}_x(\mathbf{1}_{B_1}, \mathbf{1}_{B_2}) \\ &= wp_w(1-p_w) + w(w-1)(p_{ww} - p_w^2) \\ &= w(p_w - p_{ww}) + w^2(p_{ww} - p_w^2) \\ \mathbf{Cov}_x(\tilde{X}, \tilde{Y}) &= wb(p_{wb} - p_w p_b) \\ \mathbf{Var}_x(\tilde{Y}) &= b(p_b - p_{bb}) + b^2(p_{bb} - p_b^2) \end{aligned} \tag{39}$$

Since all expectations are taken starting from x , we drop the subscript x in the proofs. By (39), all the quantities considered may be expressed in terms of p_w, p_b, p_{ww}, p_{bb} and p_{wb} . Using the asymptotic results from Theorem 4 and Lemma 28, we get the results after a short computation. For example, the last result follows from:

$$\begin{aligned} &b^2 \mathbf{Var}(\tilde{X}) - 2wb \mathbf{Cov}(\tilde{X}, \tilde{Y}) + w^2 \mathbf{Var}(\tilde{Y}) \\ &= b^2 w(p_w - p_{ww}) + b^2 w^2(p_{ww} - p_w^2) - 2w^2 b^2(p_{wb} - p_w p_b) \\ &\quad + w^2 b(p_b - p_{bb}) + w^2 b^2(p_{bb} - p_b^2) \\ &= wb \left[b(p_w - p_{ww}) + w(p_b - p_{bb}) + wb \left[p_{ww} - p_w^2 - 2p_{wb} + 2p_w p_b + p_{bb} - p_b^2 \right] \right] \\ &= wb \left[(b+w)v_s(x)(1-v_s(x)) + \mathcal{O}(1) + wb \left[p_{ww} + p_{bb} - 2p_{wb} - (p_w - p_b)^2 \right] \right] \\ &= wb \left[(b+w)v_s(x)(1-v_s(x)) + \mathcal{O}(1) \right] \end{aligned}$$

where in the last line, we have used the fact that $p_{bb} - p_{wb}$ and $p_{ww} - p_{wb}$ have the same leading term at order $1/n$, so that they cancel out.

The bounds on the higher order moments follow from the fact that \tilde{X} and \tilde{Y} are sums of negatively related indicator variables. For example, \tilde{X} is the sum of w indicators, so that by the bound (11) of Theorem 12,

$$\mathbb{E} \left[\left| \tilde{X} - \mathbb{E} [\tilde{X}] \right|^3 \right] \leq 12ew^{3/2} \leq Cn^{3/2}. \quad \square$$

4.4 Infinitesimal mean

From this moment on, until Section 4.8, the parameter β is fixed to 0. We set

$$\tilde{Z} = \frac{\tilde{X}}{\tilde{X} + \tilde{Y}} = \phi(\tilde{X}, \tilde{Y})$$

where $\phi : (x, y) \mapsto x/(x + y)$. Let \mathcal{F} be the sigma-field generated by (\tilde{X}, \tilde{Y}) . Recall that (since $\beta = 0$), X is the proportion of white balls after a binomial sampling with probability \tilde{Z} . By conditioning,

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

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

Lemma 29 (Expectation of \tilde{Z}). *The first moment of \tilde{Z} is given by*

$$\mathbb{E} [\tilde{Z}] - x = -\frac{1}{n} \cdot \frac{x(1-x)v'_s(x)}{v_s^2(x)} + \mathcal{O}(1/n^{3/2}). \quad (40)$$

Corollary 30. *The formula (34) holds when $\beta = 0$.*

Proof. The proportion \tilde{Z} is a function of \tilde{X} and \tilde{Y} . We wish to apply Taylor's formula to ϕ to compare $\mathbb{E}_x [\tilde{Z}]$ to $\phi(\mathbb{E}_x [\tilde{X}], \mathbb{E}_x [\tilde{Y}]) = \phi(\tilde{x}, \tilde{y})$. The derivatives of ϕ are given by:

$$\begin{aligned} \partial_1 \phi(x, y) &= \frac{y}{(x+y)^2}, & \partial_2 \phi(x, y) &= \frac{-x}{(x+y)^2} \\ \partial_{11} \phi(x, y) &= \frac{-2y}{(x+y)^3} & \partial_{22} \phi(x, y) &= \frac{2x}{(x+y)^3} \\ \partial_{12} \phi(x, y) &= \frac{x-y}{(x+y)^3}. \end{aligned} \quad (41)$$

Let us apply Taylor's formula to ϕ on the segment $S = [(\tilde{x}, \tilde{y}), (\tilde{X}, \tilde{Y})]$.

$$\phi(\tilde{X}, \tilde{Y}) - \phi(\tilde{x}, \tilde{y}) = T_1 + T_2 + T_3 \quad (42)$$

where

$$\begin{aligned} T_1 &= \partial_1 \phi(\tilde{x}, \tilde{y})(\tilde{X} - \tilde{x}) + \partial_2 \phi(\tilde{x}, \tilde{y})(\tilde{Y} - \tilde{y}), \\ T_2 &= \frac{1}{2} \partial_{11} \phi(\tilde{x}, \tilde{y})(\tilde{X} - \tilde{x})^2 + \partial_{12} \phi(\tilde{x}, \tilde{y})(\tilde{X} - \tilde{x})(\tilde{Y} - \tilde{y}) + \frac{1}{2} \partial_{22} \phi(\tilde{x}, \tilde{y})(\tilde{Y} - \tilde{y})^2, \end{aligned}$$

and T_3 is a remainder term which will be considered later.

We take expectations on both sides, once more dropping the subscript x from the notation. The first-order term T_1 disappears, so

$$\mathbb{E}_x [\tilde{Z}] - x = \mathbb{E} [\phi(\tilde{X}, \tilde{Y})] - x = (\phi(\tilde{x}, \tilde{y}) - x) + \mathbb{E} [T_2] + \mathbb{E} [T_3]. \quad (43)$$

Let us look at these three terms in turn. For the first one:

$$\begin{aligned} \phi(\tilde{x}, \tilde{y}) - x &= \frac{wp_w}{wp_w + bp_b} - \frac{w}{w+b} \\ &= x \left(\frac{p_w - xp_w - (1-x)p_b}{xp_w + (1-x)p_b} \right) \\ &= x(1-x) \frac{p_w - p_b}{xp_w + (1-x)p_b} \\ &= -\frac{1}{n} x(1-x) \frac{v'_s(x)}{v_s(x)} + \mathcal{O}(1/n^2), \end{aligned} \quad (44)$$

where we used Theorem 4 and the fact that $N = (1+s)n$ in the last line.

The second term T_2 is a bit trickier.

$$\mathbb{E} [T_2] = \frac{1}{2} \partial_{11} \phi(\tilde{x}, \tilde{y}) \mathbf{Var} (\tilde{X}) + \partial_{12} \phi(\tilde{x}, \tilde{y}) \mathbf{Cov} (\tilde{X}, \tilde{Y}) + \frac{1}{2} \partial_{22} \phi(\tilde{x}, \tilde{y}) \mathbf{Var} (\tilde{Y}).$$

First, remark that $\partial_{11} \phi(\tilde{x}, \tilde{y}) = \frac{1}{n^2 v_s(x)^2} \partial_{11} \phi(x, 1-x) + \mathcal{O}(1/n^3)$, and that similar results hold for the other derivatives, so that, using the rough bounds on the variances from Proposition 26, we get

$$\begin{aligned} \mathbb{E} [T_2] &= \frac{1}{n^2 v_s(x)^2} \left(\frac{1}{2} \partial_{11} \phi(x, 1-x) \mathbf{Var} (\tilde{X}) + \partial_{12} \phi(x, 1-x) \mathbf{Cov} (\tilde{X}, \tilde{Y}) \right. \\ &\quad \left. + \frac{1}{2} \partial_{22} \phi(x, 1-x) \mathbf{Var} (\tilde{Y}) \right) + \mathcal{O}(1/n^2). \end{aligned}$$

Due to the explicit expression of the derivatives (equation (41)), the term between brackets is, up to a factor n , the one that appears in Proposition 26, so

$$\mathbb{E} [T_2] = \frac{1}{n} x(1-x) \frac{v'_s(x)(v_s(x) - 1)}{v_s(x)^2} + \mathcal{O}(1/n^2). \quad (45)$$

We will prove in the next section that $T_3 = \mathcal{O}(1/n^{3/2})$. Inserting (44) and (45) in (43), we obtain (40):

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

Multiplying by n , we get (34), which proves the corollary. \square

4.5 The remainder

Let us bound the remainder term T_3 in Taylor's formula (42). If we let $z_0 = (x_0, y_0) = (\tilde{x}, \tilde{y})$ and $z_1 = (x_1, y_1) = (\tilde{X}, \tilde{Y})$ (z_0 is fixed and z_1 is random), then T_3 can be written as:

$$T_3 = \sum_{\alpha, |\alpha|=3} \frac{3}{\alpha!} \int_0^1 (1-t)^2 D_\alpha \phi(z_t) dt \cdot (z_1 - z_0)^\alpha,$$

where $z_t = (1-t)z_0 + tz_1$. To get the claimed bound on $\mathbb{E}[T_3]$, it suffices to show that, for any multi-index α of length 3,

$$R_\alpha = \mathbb{E} \left[\int_0^1 |\partial_\alpha \phi(z_t)| dt |(z_1 - z_0)^\alpha| \right] = \mathcal{O}(1/n^{3/2}).$$

Therefore we have to bound the third derivatives of ϕ on the segment $[z_0, z_1]$. The difficulty here is that (\tilde{X}, \tilde{Y}) may be very close to $(0, 0)$, where the derivatives of ϕ blow up. This problem only occurs if both \tilde{X} and \tilde{Y} are small. However, if $x \geq 1/2$, \tilde{X} is unlikely to be small, and if $x < 1/2$ then \tilde{Y} should not be too small. This prompts us to introduce the following good event:

$$A = \begin{cases} \{\tilde{X} \geq \tilde{x}/2\} & \text{if } x \geq 1/2, \\ \{\tilde{Y} \geq \tilde{y}/2\} & \text{if } x < 1/2. \end{cases} \quad (46)$$

Step 1: a bad bound on the bad event All third derivatives of ϕ satisfy:

$$|\partial_\alpha \phi(x, y)| \leq \frac{C}{(x+y)^4} (|x| + |y|).$$

Since $1 \leq \tilde{X} + \tilde{Y} \leq n$ (at least one ball is chosen), $1 \leq \tilde{x} + \tilde{y} = \mathbb{E}[\tilde{X} + \tilde{Y}] \leq n$, so that for all $t \in [0, 1]$,

$$|\partial_\alpha \phi(z_t)| \leq Cn. \quad (47)$$

This bound is not strong but it holds even on the “bad event” A^c .

Step 2: a good bound on the good event Suppose $x \geq 1/2$, so that on A , $\tilde{X} \geq \tilde{x}/2$. By the asymptotic result of Proposition 26 on \tilde{x} , and the fact that v_s is bounded below by s ,

$$\forall t, \quad |\partial_\alpha \phi(z_t)| \leq \frac{Cn}{(1/2)^4 (\tilde{x})^4} \leq \frac{C'}{n^3}.$$

If $x < 1/2$, $\tilde{Y} \geq \tilde{y}/2$ on A , so

$$\forall t, \quad |\partial_\alpha \phi(z_t)| \leq \frac{Cn}{(1/2)^4 (\tilde{y})^4} \leq \frac{C'}{n^3}.$$

Step 3: the good event has very high probability Suppose first that $x \geq 1/2$, so $A = \{\tilde{X} \geq \mathbb{E}[\tilde{X}]/2\}$. Intuitively, since $\tilde{x} = \mathbb{E}[\tilde{X}]$ is of order n and its standard deviation is of order \sqrt{n} , the event $\tilde{X} \leq \tilde{x}/2$ should have very small probability. This rigorous proof follows from the deviation bounds established above. Indeed

$$\begin{aligned} \mathbb{P}[A^c] &\leq \mathbb{P} \left[|\tilde{X} - \mathbb{E}[\tilde{X}]| > \mathbb{E}[\tilde{X}]/2 \right] \\ &\leq \exp \left(-\frac{\mathbb{E}[\tilde{X}]^2}{16xn} \right) \end{aligned}$$

where we used (12) applied to \tilde{X} , a sum of xn negatively correlated indicator. Since $x \geq 1/2$, for some absolute constant C we find that

$$\mathbb{P}[A^c] \leq \exp(-Cn). \quad (48)$$

If $x < 1/2$, $A = \{\tilde{Y} \geq \mathbb{E}[\tilde{Y}]/2\}$, and we can apply (12) to \tilde{Y} to see that (48) still holds.

Step 4: conclusion For any multiindex α , using the good bounds on the event A , and the crude bounds (47) and $|\tilde{X} - \tilde{x}| \leq 2n$, we get:

$$\begin{aligned} R_\alpha &= \mathbb{E} \left[\int_0^1 |\partial_\alpha \phi(z_t)| dt |(z_1 - z_0)^\alpha| \right] \\ &= \frac{C}{n^3} \mathbb{E} \left[\mathbf{1}_A |(\tilde{X} - \tilde{x})^\alpha| \right] + Cn^4 \exp(-Cn). \end{aligned}$$

The third moment bounds from Proposition 26 yield:

$$R_\alpha = \frac{C}{n^{3/2}} + Cn^4 \exp(-Cn) = \mathcal{O}(n^{-3/2}).$$

This proves that $\mathbb{E}[T_3] = \mathcal{O}(n^{-3/2})$, and concludes the proof of the formula (34) for the infinitesimal mean in the case $\beta = 0$.

4.6 Infinitesimal variance

Let us first study the second moment of \tilde{Z} .

Lemma 31 (Variance of \tilde{Z}). *The variance of \tilde{Z} is given by*

$$\mathbf{Var}(\tilde{Z}) = \frac{1}{n} x(1-x) \frac{v_s(x)(1-v_s(x))}{v_s(x)^2} + \mathcal{O}(1/n^{3/2}). \quad (49)$$

Proof. Since we already know the behaviour of $\mathbb{E}[\tilde{Z}]$, we only need to compute $\mathbb{E}_x[\tilde{Z}^2]$. To this end let $\psi(x, y) = \phi(x, y)^2$, so that $\mathbb{E}_x[\tilde{Z}^2] = \mathbb{E}_x[\psi(\tilde{X}, \tilde{Y})]$ and we can use Taylor's formula once again: $\psi(\tilde{X}, \tilde{Y}) - \psi(\tilde{x}, \tilde{y}) = T'_1 + T'_2 + T'_3$, where T'_i is the i^{th} order term. Taking expectations (dropping once more the subscript x) yields:

$$\mathbb{E}[\tilde{Z}^2] = \psi(\tilde{x}, \tilde{y}) + \mathbb{E}[T'_2] + \mathbb{E}[T'_3].$$

The term T'_3 is treated as before to get:

$$\mathbb{E}[T'_3] = \mathcal{O}(1/n^{3/2}).$$

To treat T'_2 we compute the derivatives of ψ :

$$\begin{aligned} \partial_1 \psi(x, y) &= \frac{2xy}{(x+y)^3}, & \partial_2 \psi(x, y) &= \frac{-2x^2}{(x+y)^3} \\ \partial_{11} \psi(x, y) &= \frac{2y(y-2x)}{(x+y)^4}, & \partial_{22} \psi(x, y) &= \frac{6x^2}{(x+y)^4}, \\ \partial_{12} \psi(x, y) &= \frac{2x(x-2y)}{(x+y)^3}. \end{aligned} \quad (50)$$

Therefore $\mathbb{E}[T'_2]$ is given by:

$$\mathbb{E}[T'_2] = \frac{1}{2} \partial_{11} \psi(\tilde{x}, \tilde{y}) \mathbf{Var}(\tilde{X}) + \partial_{12} \psi(\tilde{x}, \tilde{y}) \mathbf{Cov}(\tilde{X}, \tilde{Y}) + \frac{1}{2} \partial_{22} \psi(\tilde{x}, \tilde{y}) \mathbf{Var}(\tilde{Y}).$$

As before, we can approximate the derivatives at (\tilde{x}, \tilde{y}) by the ones at (x, y) since

$$\left| \partial_{11} \psi(\tilde{x}, \tilde{y}) - \frac{1}{n^2 v_s(x)^2} \partial_{11} \psi(x, 1-x) \right| \leq \mathcal{O}(1/n^3).$$

Using the expressions of the partial derivatives of ψ , and rearranging terms to use the results of Proposition 26, we get:

$$\begin{aligned}
\mathbb{E} [T_2'] &= \frac{1}{n^2 v_s(x)^2} \left(\frac{1}{2} (2y(y-2x)) \mathbf{Var}(\tilde{X}) + (2x(x-2y)) \mathbf{Cov}(\tilde{X}, \tilde{Y}) + \frac{1}{2} 6x^2 \mathbf{Var}(\tilde{Y}) \right) \\
&= \frac{2x}{n^2 v_s(x)^2} \left(-y \mathbf{Var}(\tilde{X}) + (x-y) \mathbf{Cov}(\tilde{X}, \tilde{Y}) + x \mathbf{Var}(\tilde{Y}) \right) \\
&\quad + \frac{1}{n^2 v_s(x)^2} \left(y^2 \mathbf{Var}(\tilde{X}) - 2xy \mathbf{Cov}(\tilde{X}, \tilde{Y}) + x^2 \mathbf{Var}(\tilde{Y}) \right) \\
&= \frac{x(1-x)}{n v_s(x)^2} \left(2xv_s'(x)(v_s(x)-1) + v_s(x)(1-v_s(x)) + \mathcal{O}(1/n) \right).
\end{aligned}$$

The last step is to find the behaviour of $\psi(\tilde{x}, \tilde{y})$. Once more, by Taylor's formula:

$$\begin{aligned}
\psi(\tilde{x}, \tilde{y}) &= \psi\left(x, (1-x) \frac{p_b}{p_w}\right) \\
&= \psi\left(x, (1-x) \left(1 + \frac{p_b - p_w}{p_w}\right)\right) \\
&= \psi(x, (1-x)) + \partial_2 \psi(x, 1-x) \cdot (1-x) \frac{v_s'(x)}{n v_s(x)} + \mathcal{O}(1/n^2) \\
&= x^2 - \frac{1}{n} 2x^2(1-x) \frac{v_s'(x)}{v_s(x)} + \mathcal{O}(1/n^2).
\end{aligned}$$

Therefore:

$$\mathbb{E} [\tilde{Z}^2] = x^2 + \frac{x(1-x)}{n v_s(x)^2} \left(-2xv_s'(x) + v_s(x)(1-v_s(x)) + \mathcal{O}(1/n) \right).$$

Since, by (40),

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

we finally obtain the expression (49) for the variance of \tilde{Z} . \square

Proof of the formula (33) for the infinitesimal variance. Since $\beta = 0$, the conditional law of nX given \mathcal{F} is the binomial law $\mathcal{B}(n, \tilde{Z})$. The variance of X is given by conditioning on \mathcal{F} :

$$\begin{aligned}
\mathbf{Var}_x(X) &= \mathbb{E}_x [\mathbf{Var}_x(X|\mathcal{F})] + \mathbf{Var}_x(\mathbb{E}_x[X|\mathcal{F}]) \\
&= \mathbb{E}_x \left[\frac{1}{n} \tilde{Z}(1-\tilde{Z}) \right] + \mathbf{Var}_x(\tilde{Z}).
\end{aligned}$$

We rearrange terms on the right hand side to get:

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

The asymptotics of $\mathbb{E}[\tilde{Z}]$ and $\mathbf{Var}_x(\tilde{Z})$ are known from lemmas 29 and 31. Injecting them in the last equation we get

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

The infinitesimal variance is obtained by multiplying by n :

$$\begin{aligned} a_n(x) &= x(1-x)\frac{v_s(x)(1-v_s(x))}{v_s^2(x)} + x(1-x) + \mathcal{O}(1/n^{1/2}) \\ &= \frac{1}{v_s(x)}x(1-x) + \mathcal{O}(1/n^{1/2}). \end{aligned} \quad \square$$

4.7 No jumps at the limit

We have to show that $n\mathbb{E}_x[|X-x|^3] \rightarrow 0$. Recalling that $\tilde{Z} = \phi(\tilde{X}, \tilde{Y}) = \tilde{X}/(\tilde{X} + \tilde{Y})$, and using the trivial bound $(a+b+c)^3 \leq 9(a^3+b^3+c^3)$,

$$n\mathbb{E}_x[|X-x|^3] \leq 9n\mathbb{E}_x\left[|X-\tilde{Z}|^3\right] + 9n\mathbb{E}_x\left[\left|\tilde{Z}-\frac{\tilde{x}}{\tilde{x}+\tilde{y}}\right|^3\right] + 9n\left(\frac{\tilde{x}}{\tilde{x}+\tilde{y}}-x\right)^3. \quad (51)$$

For the first term, we condition by the first step:

$$\mathbb{E}_x\left[|X-\tilde{Z}|^3\right] = \mathbb{E}_x\left[\mathbb{E}\left[|X-\tilde{Z}|^3\middle|\mathcal{F}\right]\right].$$

Given \mathcal{F} , nX follows a binomial law of parameters xn and \tilde{Z} . Using (for example) the bound (11) (which holds in the more general negatively dependent case), the whole term is $\mathcal{O}(n^{-1/2})$.

The third term of (51) is $\mathcal{O}(1/n^2)$, thanks to the controls on \tilde{x} and \tilde{y} from Proposition 26.

Therefore we only have to bound the second term $n\mathbb{E}_x\left[\left|\tilde{Z}-\frac{\tilde{x}}{\tilde{x}+\tilde{y}}\right|^3\right]$. Let us reuse the notation $z_t = (1-t)(\tilde{x}, \tilde{y}) + t(\tilde{X}, \tilde{Y})$:

$$n\mathbb{E}_x\left[\left|\tilde{Z}-\frac{\tilde{x}}{\tilde{x}+\tilde{y}}\right|^3\right] = n\mathbb{E}_x\left[|\phi(z_1) - \phi(z_0)|^3\right].$$

As in Section 4.5, we want to bound the derivatives of ϕ on the segment $[z_0, z_1]$, which is only possible if $\tilde{X} + \tilde{Y}$ is large enough. Recall the good event A from Equation (46). On A , it is easy to see that all first derivatives of ϕ are bounded by C/n for some absolute constant C , therefore

$$n\mathbb{E}_x\left[\left|\tilde{Z}-\frac{\tilde{x}}{\tilde{x}+\tilde{y}}\right|^3\right] \leq \frac{C}{n^2}\left(\mathbb{E}\left[(\tilde{X}-\tilde{x})^3\right] + \mathbb{E}\left[(\tilde{Y}-\tilde{y})^3\right]\right) + C\mathbb{P}[A^c]n^4.$$

The third moments are controlled by Proposition 26, and the probability of the bad event is exponentially small (by (48)). This shows that (35) holds, and concludes the proof of the main result when $\beta = 0$.

4.8 Proofs for the full model

In this final section we show how to compute the infinitesimal mean and variance in the general case, that is, we prove Proposition 23 when β 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 σ -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}_β . This modified ratio is not far from \tilde{Z} :

$$\begin{aligned} \tilde{Z}_\beta &= \frac{(1 + \beta/n)\tilde{X}}{(1 + \beta/n)\tilde{X} + \tilde{Y}} = \tilde{Z} \left(\frac{1 + \beta/n}{(1 + \beta/n)\tilde{Z} + 1 - \tilde{Z}} \right) \\ &= \tilde{Z}(1 + \beta/n)(1 - (\beta/n)\tilde{Z} + \mathcal{O}(1/n^2)) \\ &= \tilde{Z} \left(1 + (\beta/n)(1 - \tilde{Z}) + \mathcal{O}(1/n^2) \right), \end{aligned} \tag{52}$$

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

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

Now let us recall the results from lemmas 29 and 31:

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

This immediately entails

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

and proves the general form of the infinitesimal mean announced in Proposition 23.

For the variance, squaring (52) and taking expectations gives

$$\mathbb{E} [\tilde{Z}_\beta^2] = \mathbb{E} [\tilde{Z}^2] + 2(\beta/n)\mathbb{E} [\tilde{Z}^2(1 - \tilde{Z})] + \mathcal{O}(1/n^2),$$

therefore

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

The absolute value of the second term is bounded above by $\frac{2\beta}{n} \mathbf{Var} (\tilde{Z})^{1/2}$; since $\mathbf{Var} (\tilde{Z})$ is of order $1/n$ we get

$$\mathbf{Var} (\tilde{Z}_\beta) = \mathbf{Var} (\tilde{Z}) + \mathcal{O}(1/n^{3/2}),$$

which proves that the infinitesimal variance does not depend on β .

Finally we have to show that the control on the jump sizes still holds. Adding one more intermediate term in 51 we get:

$$\begin{aligned} n\mathbb{E} \left[|X - x|^3 \right] \leq & 16n \left(\mathbb{E} \left[|X - \tilde{Z}_\beta|^3 \right] + \mathbb{E} \left[|\tilde{Z}_\beta - \tilde{Z}|^3 \right] \right. \\ & \left. + \mathbb{E} \left[\left| \tilde{Z} - \frac{\tilde{x}}{\tilde{x} + \tilde{y}} \right|^3 \right] + \left(\frac{\tilde{x}}{\tilde{x} + \tilde{y}} - x \right)^3 \right) \end{aligned}$$

The first, third and fourth terms are treated exactly as in Section 4.7, since \tilde{Z}_β is the conditional expectation of X knowing \mathcal{F} . For the second term, recalling (52), we get

$$\left| \tilde{Z}_\beta - \tilde{Z} \right|^3 = \frac{\beta^3}{n^3} \left(\tilde{Z}(1 - \tilde{Z}) + \mathcal{O}(1/n) \right)^3,$$

which implies that $n\mathbb{E} \left[\left| \tilde{Z} - \tilde{Z}_\beta \right|^3 \right]$ converges to zero. This concludes the proof of Proposition 23 in the general case.

Acknowledgements. This paper 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. Couplier for interesting discussions. Finally we thank J.E. Taylor and an anonymous reviewer for many valuable comments on a preliminary version of this work.

References

- [BHJ92] A. D. Barbour, Lars Holst, and Svante Janson, *Poisson approximation*, Oxford Studies in Probability, vol. 2, The Clarendon Press Oxford University Press, New York, 1992, Oxford Science Publications. MR 1163825 (93g:60043)
- [CFM06] N. Champagnat, R. Ferriere, and S. Meleard, *Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models*, Theor Popul Biol **69** (2006), no. 3, 297–321.
- [CFM08] Nicolas Champagnat, Régis Ferrière, and Sylvie Méléard, *From individual stochastic processes to macroscopic models in adaptive evolution*, Stoch. Models **24** (2008), no. suppl. 1, 2–44. MR 2466448 (2010b:60238)
- [CS09] Fabio A. C. C. Chalub and Max O. Souza, *A non-standard evolution problem arising in population genetics*, Commun. Math. Sci. **7** (2009), no. 2, 489–502. MR 2536449 (2010h:92102)
- [CS14] ———, *The frequency-dependent Wright-Fisher model: diffusive and non-diffusive approximations*, J. Math. Biol. **68** (2014), no. 5, 1089–1133. MR 3175200

- [DB06] David Damiens and Guy Boivin, *Why do sperm-depleted parasitoid males continue to mate?*, Behavioral Ecology **17** (2006), no. 1, 138–143.
- [Dio07] Francisco Dionisio, *Selfish and spiteful behaviour through parasites and pathogens.*, Evol. Ecol. Res. **9** (2007), no. 7, 1199–1210.
- [Dur96] Richard Durrett, *Stochastic calculus: a practical introduction*, Probability and Stochastics Series, CRC PressINC, 1996.
- [Dur08] ———, *Probability models for DNA sequence evolution*, second ed., Probability and its Applications (New York), Springer, New York, 2008. MR 2439767 (2009k:60001)
- [EK86] Stewart N. Ethier and Thomas G. Kurtz, *Markov processes*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1986, Characterization and convergence. MR 838085 (88a:60130)
- [Eth11] Alison Etheridge, *Some mathematical models from population genetics*, Lecture Notes in Mathematics, vol. 2012, Springer, Heidelberg, 2011, Lectures from the 39th Probability Summer School held in Saint-Flour, 2009. MR 2759587 (2011j:92043)
- [Ewe04] Warren J. Ewens, *Mathematical population genetics. I*, second ed., Interdisciplinary Applied Mathematics, vol. 27, Springer-Verlag, New York, 2004, Theoretical introduction. MR 2026891 (2004k:92001)
- [FWR01] K. R. Foster, T. Wenseleers, and F. L. W. Ratnieks, *Spite: Hamilton's unproven theory.*, Annales Zoologici Fennici **38** (2001), no. 3-4, 229–238.
- [Gil74] J. H. Gillespie, *Natural selection for within-generation variance in offspring number*, Genetics **76** (1974), no. 3, 601–606.
- [Gil75] ———, *Natural selection for within-generation variance in offspring number II. Discrete haploid models*, Genetics **81** (1975), no. 2, 403–413.
- [GS01] Geoffrey R. Grimmett and David R. Stirzaker, *Probability and random processes*, third ed., Oxford University Press, New York, 2001. MR 2059709 (2004m:60002)
- [Ham70] W. D. Hamilton, *Selfish and spiteful behaviour in an evolutionary model*, Nature **228** (1970), no. 5277, 1218–1220.
- [Jan94] Svante Janson, *Large deviation inequalities for sums of indicator variables*, 1994.
- [JDP83] Kumar Joag-Dev and Frank Proschan, *Negative association of random variables, with applications*, Ann. Statist. **11** (1983), no. 1, 286–295. MR 684886 (85d:62058)
- [Les05] Sabin Lessard, *Long-term stability from fixation probabilities in finite populations: New perspectives for ESS theory.*, Theoret. Popul. Biology **68** (2005), no. 1, 19–27.

- [MW07] A. J. McKane and D. Waxman, *Singular solutions of the diffusion equation of population genetics*, J. Theoret. Biol. **247** (2007), no. 4, 849–858. MR 2479629 (2009m:92082)
- [Ric96] W. R. Rice, *Sexually antagonistic male adaptation triggered by experimental arrest of female evolution*, Nature **381** (1996), no. 6579, 232–234.
- [Ros11] Nathan Ross, *Fundamentals of Stein’s method*, Probab. Surv. **8** (2011), 210–293. MR 2861132 (2012k:60079)
- [RPSWT09] Preethi Radhakrishnan, Diana Pérez-Staples, Christopher W. Weldon, and Phillip W. Taylor, *Multiple mating and sperm depletion in male queensland fruit flies: effects on female remating behaviour*, Animal Behaviour **78** (2009), no. 4, 839–846.
- [Shp07] M. Shpak, *Selection against demographic stochasticity in age-structured populations*, Genetics **177** (2007), no. 4, 2181–2194.
- [SHR08] Sven Steiner, Nina Henrich, and Joachim Ruther, *Mating with sperm-depleted males does not increase female mating frequency in the parasitoid *Lariophagus distinguendus**, Entomologia Experimentalis et Applicata **126** (2008), no. 2, 131–137.
- [Tay09] J. E. Taylor, *The genealogical consequences of fecundity variance polymorphism*, Genetics **182** (2009), no. 3, 813–837.
- [Wax11] 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** (2011), 79–87. MR 2974468

Ludovic GOUDENÈGE, e-mail: [goudenege\(AT\)math.cnrs.fr](mailto:goudenege(AT)math.cnrs.fr)

Pierre-André ZITT, e-mail: [pierre-andre.zitt\(AT\)u-pem.fr](mailto:pierre-andre.zitt(AT)u-pem.fr)

LAMA, UNIVERSITÉ DE MARNE-LA-VALLÉE, 5, BOULEVARD DESCARTES, CITÉ DESCARTES - CHAMPS-SUR-MARNE, 77454 MARNE-LA-VALLÉE CEDEX 2, FRANCE.