The rate of the Muller's ratchet.

Anton Goloborodko.

05/18/2013

1 Introduction.

The majority of new mutations arising in well-adapted species have a negative effect on fitness [1]. Sexual species get rid of the damaged copies of genes by recombination between homologous chromosomes, effectively sharing the best available versions of genes with each other. In asexual populations though, the population consists of a number of non-interacting lineages, each steadily accumulating their own mutations and passing them down to next generations. Supported by stochastic effects, this process may lead to extinction of the best-adapted genomes from the population. This process was first described by Hermann Joseph Muller in 1932 and was later dubbed the "Muller's ratchet", emphasizing its irreversibility. Muller's ratchet was observed both in experiments and in nature [2, 3, 4, 5]. A recent work [6] showed that this process to play an important role in the dynamics of cancer.

Despite the long history of the subject, our understanding of the Muller's ratchet is incomplete. Many efforts have been done to estimate the rate of mutation accumulation [7, 8, 9, 10], including very recent ones [11, 12, 13]. Here we review these efforts and compile them into a complete solution accessible to people outside of the field. In section 2, we start with the basic mathematical formulation of population genetics and derive the properties of the Wright-Fisher model. A reader familiar with the stochastic formulation of population genetics and the conception of the mutation-selection balance may skip that section. In section 3 we review and present the previously described mathematical models of the Muller's ratchet.

2 Wright-Fisher model.

The Wright-Fisher model is the simplest and most popular theoretical tool to describe evolving populations. The model considers a population of N individuals. At each generation, individuals are chosen randomly to produce N children, after which the parental generation dies out. The numbers of progeny per parent are approximately independent and Poisson-distributed with an average of 1.

The individuals have their own genomes that they pass to their progeny. The genome defines the fitness f of its owner, the quantitative measure of how well the individual is adapted to the environment. The fitness defines the expected number of progeny as $\frac{f}{\bar{f}}$, where $\bar{f}$ is the average fitness across the population.

2.1 Kolmogorov's equations.

The WF model is inherently stochastic due to random sampling of parents at each generation. A powerful approach to describe the time evolution of random systems is to use the continuous time and state density functions $f(x;p,t)$, the probability of the system to come into a state x time t after it started from a state p. Then the time evolution of $f(x;p,t)$ will approximately obey the following equations:

$$\frac{\partial f(x;p,t)}{\partial t} = -\frac{\partial}{\partial x}\left[a(x)f(x;p,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial^2 x}\left[b(x)f(x;p,t)\right] \tag{1}$$

$$\frac{\partial f(x;p,t)}{\partial t} = a(p)\frac{\partial f(x;p,t)}{\partial p} + \frac{1}{2}b(p)\frac{\partial^2 f(x;p,t)}{\partial^2 p} \tag{2}$$

where functions $a(x)$ and $b(x)$ are given by the expressions:

$$a(x) = \frac{\langle \delta x \rangle (x)}{\delta t} \tag{3}$$

$$b(x) = \frac{\left\langle \delta x^2 - \left\langle \delta x^2 \right\rangle \right\rangle (x)}{\delta t} \tag{4}$$

or, in simpler terms describe the expected size and standard deviation of a step δx that a system makes from a state x over time δt. Equations (1) and (2) are called forward and backward Kolmogorov equations, correspondingly, and formally correct if the system doesn't make distant steps in short periods of time. Note that the forward Kolmogorov equation (1) has a very intuitive interpretation: the probability density "flows" between different states in a directed and random fashions, proportional to $f(x)$ and its first derivative $\frac{\partial f(x)}{\partial x}$, correspondingly, with coefficients of proportionality varying between states. The change of probability density at a particular state x is caused be the local difference between the incoming and outgoing probability currents expressed by an extra derivative $\frac{\partial}{\partial x}$.

2.2 Single gene systems.

Let us study a simple case of a population of N individuals, each having one gene with two states (alleles) with fitnesses 1 and $1+s$, $s \ll 1$ [14, 15, 16]. If the portion of individuals having the better allele is x, then in the next generation their portion x' will be Poisson-distributed:

$$p(x';x,s,N)=\binom{x'N}{N}\left(xN\frac{1+s}{\bar{f}}\right)^{x'N}\left((1-x)N\frac{1}{\bar{f}}\right)^{(1-x')N}=$$

$$=\binom{x'N}{N}\left(x(1+s)\right)^{x'N}(1-x)^{(1-x')N}\frac{1}{(1+xs)^N}$$

Applying (3) and (4) to this expression with $x'=x+\delta x$ and $\delta t=1$ gives us:

$$a(x)=sx(1-x) \tag{5}$$

$$b(x)=\frac{x(1-x)}{N} \tag{6}$$

Combined with the Kolmogorov equations (1) and (2) this fully describes the dynamics of the one gene system.

Note that both $a(x)$ and $b(x)$ turn to zero at $x=0$ and $x=1$. This reflect the fundamental property of the one-gene system in the absence of mutations: once one of the gene alleles goes extinct, the system comes into a steady state. Moreover, at infinite time the system has to go to either of the two steady states. We could apply the obtained relationships (5) and (6) to find the probability of the better allele to become extinct $P_0(p)$ or take over the whole population $P_1(p)$ starting from the initial frequency p.

Setting $x=0$ in (2) gives us the differential equation for $P_0(t,p)$:

$$\frac{\partial P_0(p,t)}{\partial t}=a(p)\frac{\partial P_0(p,t)}{\partial p}+\frac{1}{2}b(p)\frac{\partial^2 P_0(p,t)}{\partial^2 p} \tag{7}$$

At $t\to\infty$, $P_0(t,p)\to P_0(p)$ and $\frac{\partial}{\partial t}P_0(t,p)\to 0$,

$$0=a(p)\frac{\partial P_0(p)}{\partial p}+\frac{1}{2}b(p)\frac{\partial^2 P_0(p)}{\partial^2 p} \tag{8}$$

$$\frac{P_0''(p)}{P_0'(p)}=\frac{d}{dp}\log P_0'(p)=-\frac{2a(p)}{b(p)}$$

$$\log P_0'(p)=\log P_0'(p)-\int\limits_0^p\frac{2a(p)}{b(p)}dp$$

$$P_0'(p)=P_0'(0)\,exp\left(-\int\limits_0^p\frac{2a(p)}{b(p)}dp\right)=P_0'(0)s(p)$$

where

$$s(p)=exp\left(-\int\limits_0^p\frac{2a(p)}{b(p)}dp\right) \tag{9}$$

A boundary condition on $P_0(p)$ comes from the fact that an allele cannot go extinct after it reaches 100% frequency, $P_0(1) = 0$:

$$P_0(p) = P_0(1) - \int\limits_1^p P_0'(0)s(p')dp' = \int\limits_p^1 P_0'(0)s(p')dp'$$

$$P_0(0) = 1, \Longrightarrow P_0'(0) = \frac{1}{\int_0^1 s(p')dp'}$$

The resulting formulae for $P_0(p)$ and $P_1(p)$ can be nicely expressed using the scale function $S(x)$ [16]:

$$S(p) = \int\limits^p s(p')dp'$$

where the lower limit of integration is chosen arbitrarily, but never enters the resulting expressions:

$$P_0(p) = \frac{S(p) - S(0)}{S(1) - S(0)}$$

$$P_1(p) = \frac{S(1) - S(p)}{S(1) - S(0)}$$

Which gives us:

$$P_1(p) = \frac{1 - e^{-2Nsp}}{1 - e^{-2Ns}}$$

$$P_0(p) = 1 - P_1(p)$$

2.3 Many-gene systems. Mutation-selection balance.

The number of genes in real organisms varies from hundreds to tens of thousands. Every gene affects its owner in a different way and, to further complicate things, the fitness gain from a gene may depend on the presence of other genes. However, even the simplest models give useful insights about the real systems.

Let us consider a model where individuals have an infinite number of two-state genes. Unlike in the previous subsection, here the genes can be in a most-fit or a mutated state, with individual's fitness f depending on the number of its mutated genes k as $f = 1 - ks$. We introduce mutations as a small probability u that the genome of a progeny will obtain one extra mutation compared to the parental genome. The indistinguishability of genes allows us to split the individuals into groups according to the number of mutations they have. These groups are called mutational classes and their relative abundance is denoted by $X(m,t)$, where m is the number of mutations. The following expression captures their time evolution:

$$NX(m,t{+}1) \approx NX(m,t){+}N\frac{dX(m,t)}{dt} = NX(m,t)\frac{1-sm}{1-s\bar{m}}{-}uNX(m,t){+}uNX(m{-}1,t)$$

$$\frac{dX(m,t)}{dt} = X(m,t)\frac{s(\bar{m}-m)}{1-s\bar{m}} - uX(m,t) + uX(m-1,t) \tag{10}$$

$$\frac{dX(m,t)}{dt} \approx X(m,t)s(\bar{m}-m) - uX(m,t) + uX(m-1,t) \tag{11}$$

The terms in this expression correspond to the change in the frequency due to natural selection, loss due to mutation into higher mutational classes and gain due to mutations from lower mutational classes. Note, that the equation for $X(0,t)$ does not have the last term.

In a steady state:

$$0 = \frac{dX(0,t)}{dt} \approx X(0,t)s(\bar{m}-0) - uX(0,t)$$

Therefore,

$$\bar{m} = \frac{u}{s}$$

$$0 = \frac{dX(m,t)}{dt} = X(m)s\left(\frac{u}{s} - m\right) - uX(m) + uX(m-1)$$

$$X(m) = \frac{u}{ms}X(m-1)$$

The above relationship produces a Poisson distribution of the frequency of mutational classes:

$$X(m) = \frac{1}{m!}\left(\frac{u}{s}\right)^m e^{-\frac{u}{s}},\ m = 0,1,2,... \tag{12}$$

An important result of the above derivation is the frequency of non-mutated genomes, or of the most-fit class in the population:

$$x = X(0) = e^{-\frac{u}{s}}$$

3 The mathematical model of the Muller's ratchet.

3.1 Stochastic diffusion and the average time to extinction.

The conceptions introduced in the previous chapter now allow us to quantitatively describe the Muller's ratchet. We begin with noting that the mutational class frequencies $X(m)$ given by (12) are the equilibrium values. The real distribution fluctuates over time around these points. However, one can imagine the situation when the fluctuations would reduce the frequency of the most fit class x to 0. Once this happens, the Muller's ratchet clicks, the population loses its intact genome and the best genome it has now is a genome with one mutation. Due to the linearity of fitness with respect to the number of mutations, the new equilibrium will again be described by the equation (12), now shifted by one mutation:

$$X(m) = \frac{1}{(m-1)!}\left(\frac{u}{s}\right)^{m-1} e^{-\frac{u}{s}},\ m = 1, 2, 3, ...$$

The stochastic loss of the most fit class is irreversible and leads to steady accumulation of mutations in the population. We can now describe the fluctuations of x using the same set of mathematical tools introduced in section 2. We will spend the rest of this section to find the general expression for $\bar{t}$(p), the average time to extinction of the most fit class, given its initial frequency p. This time is expressed as:

$$\bar{t}(p) = \int\limits_0^\infty t\phi(p,t)\,dt \tag{13}$$

where $\phi(p,t) = \frac{dP_0(p,t)}{dt}$ and equals the probability of the system to hit $x_0 = 0$ exactly at time t. The differential equation describing evolution of $\phi(p,t)$ can be obtained from (7) using time differentiation:

$$\frac{\partial\phi(p,t)}{\partial t} = a(p)\frac{\partial\phi(p,t)}{\partial p} + \frac{1}{2}b(p)\frac{\partial^2\phi(p,t)}{\partial^2 p}$$

Noting that the system inevitably hits $x_0 = 0$ at $t \to \inf$ and using partial integration, we obtain:

$$1 = \int\limits_0^\infty \phi(p,t)\,dt = t\phi(p,t)\Big|_0^\infty - \int\limits_0^\infty t\frac{\partial\phi(p,t)}{\partial t}\,dt =$$

$$= 0 - \int\limits_0^\infty t\left(a(p)\frac{\partial\phi(p,t)}{\partial p} + \frac{1}{2}b(p)\frac{\partial^2\phi(p,t)}{\partial^2 p}\right)\,dt$$

Using the definition of $\bar{t}(p)$ (13), we obtain:

$$-1 = a(p)\frac{\partial\bar{t}(p)}{\partial p} + \frac{1}{2}b(p)\frac{\partial^2\bar{t}(p)}{\partial^2 p}$$

The next non-trivial step is to recall the definition of $s(p)$ (9) and combine it with $\bar{t}'(p)$:

$$\begin{aligned}\frac{d}{dp}\left(\frac{\bar{t}'(p)}{s(p)}\right) &= \frac{d}{dp}\left(exp\left(\int^{p}\frac{2a(p')}{b(p')}dp'\right)\bar{t}'(p)\right)\\ &= exp\left(\int^{p}\frac{2a(p')}{b(p')}dp'\right)\left(\frac{2a(p)}{b(p)}\bar{t}'(p)+\bar{t}''(p)\right)\\ &= -\frac{2}{b(p)s(p)}\end{aligned}$$

Introducing the speed density function $m(p)$, we could further simplify the equation for $\bar{t}'(p)$:

$$m(p) = \frac{1}{b(p)s(p)} \tag{14}$$

$$\frac{d}{dp}\left(\frac{\bar{t}'(p)}{s(p)}\right) = -2m(p)$$

$$\frac{\bar{t}'(p)}{s(p)} = -2\int_{0}^{p} m(p')dp' + \beta \tag{15}$$

$$\bar{t}(p) = -2\int_{0}^{p} s(\xi)d\xi \int_{0}^{\xi} m(p')dp' + \beta\left(S(p) - S(0)\right) + \alpha$$

The particular form of the solution depends on the nature of boundaries. In our case, the ratchet clicks when the frequency of the most fit class x hits 0, which leads to the following boundary condition on $\bar{t}(p)$:

$$\bar{t}(0) = 0, \Longrightarrow \alpha = 0$$

The second boundary condition comes from the fact that the frequency of the most fit class cannot go above $x = 1$. This type of boundary is called a reflective one and gives an additional constraint on $\bar{t}'(p)$:

$$\bar{t}'(1) = 0, \Longrightarrow \beta = 2\int_{0}^{1} m(p')dp'$$

Here we used (15) to calculate the value of β.

$$\bar{t}(p) = -2\int_{0}^{p} s(\xi)d\xi \int_{0}^{\xi} m(p')dp' + 2\left(S(p) - S(0)\right)\int_{0}^{1} m(p')dp'$$

Reversing the order of integration:

$$\bar{t}(p) = -2\int\limits_0^p m(p')dp' \int\limits_{p'}^p s(\xi)d\xi + 2\left(S(p)-S(0)\right)\int\limits_0^p m(p')dp' + 2\left(S(p)-S(0)\right)\int\limits_p^1 m(p')dp' =$$

$$= 2\int\limits_0^p \left(-S(p)+S(p')+S(p)-S(0)\right) m(p')dp' + 2\left(S(p)-S(0)\right)\int\limits_p^1 m(p')dp'$$

The final solution for $\bar{t}(p)$ is:

$$\bar{t}(p) = 2\int\limits_0^p \left(S(p')-S(0)\right) m(p')dp' + 2\int\limits_p^1 \left(S(p)-S(0)\right) m(p')dp' \tag{16}$$

This expression also admits another form:

$$\bar{t}(p) = 2\int\limits_0^1 \left(S(p')-S(0)\right) m(p')dp' - 2\int\limits_p^1 \left(S(p')-S(p)\right) m(p')dp' \tag{17}$$

3.2 Extinction of the most fit class.

In order to estimate the average time to extinction of the most fit class using (16), we need to know the functions $a(x)$ and $b(x)$ characterizing the fluctuations of the size of the most fit class. This analysis was first done in [7] and later extended in [8] and [9]. Starting from the definition of $a(x)$ (3) and considering small perturbations around the equilibrium value $x = x_0 = e^{-\frac{u}{s}}$ we get:

$$a(x) = \frac{\langle \delta x \rangle (x)}{\delta t} = x\frac{s\bar{m}}{1-s\bar{m}} - ux = x\frac{s\left(\bar{m}-\frac{u}{s}\right)}{1-s\bar{m}} = x\frac{\delta f}{\bar{f}}$$

where we used (10) to describe the time evolution of $x(t)$. Using basic reasoning and the linear response theory [17], it was suggested in [8] that:

$$\delta f \approx \gamma s\bar{f}\left(1-\frac{x}{x_0}\right)$$

where $\gamma{\approx}0.6$ is a numerical coefficient describing the dynamics of the system right after the loss of the most fit class. Combined together this gives us:

$$a(x) = \gamma s x\left(1-\frac{x}{x_0}\right)$$

The diffusion coefficient $b(x)$ corresponds to the binomial sampling and at low x is approximated by:

$$b(x) \approx \frac{x}{N}$$

Finally, the system is thought to start at the equilibrium value of x:

$$p = x_0 = e^{-\frac{u}{s}}$$

Taken together, these expressions give us the equation for $s(x)$:

$$s(x) = exp\Bigg(-2\int_0^x \frac{\gamma s x'\left(1-\frac{x'}{x_0}\right)N}{x'}dx'\Bigg) = exp\Big(-\tfrac{2\gamma sN}{x_0}\int_0^x (x_0 - x')dx'\Big) =$$

$$= exp\left(\frac{2\gamma sN}{x_0}x\left(\frac{x^2}{2} - x_0\right)\right)$$

And the final expression for T_{click} is given by (16).

Importantly, the proposed solution requires the number of individuals in the most fit class $NX(0) = Ne^{-\frac{u}{s}} > 1$. Otherwise, the continuous description of the population using the frequencies of mutational classes $X(m)$ breaks down and the derived expression are no longer valid.

A correction to the above value of T_{click} was suggested in [10]. The authors pointed out that right after the loss of the most fit class, the frequency of the new most fit class is found away from its equilibrium value. Therefore, the total time between two clicks should be $T_{click} + T_{equilibration}$, where

$$T_{equilibration} = \frac{1}{s}\left(1 - \frac{es}{(e-1)u}\right)$$

3.3 Jain's formula.

A simplification of the integral 16 was suggested in [12]. The author discarded the second term from the expression 17 for T_{click} and showed that this expression can be further simplified to the following form:

$$T_{click} = \sqrt{\frac{\pi}{\gamma^3 s^3 N e^{-\frac{u}{s}}}}\, e^{\gamma s N e^{-\frac{u}{s}}}$$

4 Conclusions.

Here we introduced the basics stochastic models of population genetics. We used these approached to explicitly derive the most widely used estimates of the rate of Muller's ratchet. Further directions involve comparison of various approximations for the rate of the Muller's ratchet with the results of stochastic simulations.

The author thanks Christopher McFarland and Kirill Korolev for their help and productive discussions.

References

[1] Adam Eyre-Walker and Peter D Keightley. The distribution of fitness effects of new mutations. *Nature reviews. Genetics*, 8(8):610–8, August 2007.

[2] W. Rice. Degeneration of a nonrecombining chromosome. *Science*, 263(5144):230–232, January 1994.

[3] D I Andersson and D Hughes. Muller's ratchet decreases fitness of a DNA-based microbe. *Proceedings of the National Academy of Sciences of the United States of America*, 93(2):906–7, January 1996.

[4] C Zeyl, M Mizesko, and J a de Visser. Mutational meltdown in laboratory yeast populations. *Evolution; international journal of organic evolution*, 55(5):909–17, May 2001.

[5] Dana K Howe and Dee R Denver. Muller's Ratchet and compensatory mutation in Caenorhabditis briggsae mitochondrial genome evolution. *BMC evolutionary biology*, 8:62, January 2008.

[6] Christopher D McFarland, Kirill S Korolev, Gregory V Kryukov, Shamil R Sunyaev, and Leonid a Mirny. Impact of deleterious passenger mutations on cancer progression. *Proceedings of the National Academy of Sciences of the United States of America*, 110(8):2910–5, February 2013.

[7] John Haigh. The accumulation of deleterious genes in a population—Muller's Ratchet. *Theoretical Population Biology*, 14(2):251–267, October 1978.

[8] W Stephan, L Chao, and J G Smale. The advance of Muller's ratchet in a haploid asexual population: approximate solutions based on diffusion theory. *Genetical research*, 61(3):225–31, June 1993.

[9] Isabel Gordo and Brian Charlesworth. The Degeneration of Asexual Haploid Populations and the Speed of Muller ' s Ratchet. 2000.

[10] Isabel Gordo and Brian Charlesworth. Letter to the Editor On the Speed of Muller ' s Ratchet. 2140(December):2137–2140, 2000.

[11] A. Etheridge, P. Pfaffelhuber, and A. Wakolbinger. How often does the ratchet click? Facts, heuristics, asymptotics. 70, September 2007.

[12] Kavita Jain. Loss of least-loaded class in asexual populations due to drift and epistasis. *Genetics*, 179(4):2125–34, August 2008.

[13] Richard a Neher and Boris I Shraiman. Fluctuations of fitness distributions and the rate of Muller's ratchet. *Genetics*, 191(4):1283–93, August 2012.

[14] A. Etheridge. Lecture notes on diffusion process models in mathematical genetics.

[15] W.J. Ewens. *Mathematical Population Genetics: I. Theoretical Introduction.* Springer, 2004.

[16] S Karlin and H M Taylor. *A second course in stochastic processes.* Number v. 2. ACADEMIC PressINC, 1981.

[17] H. Risken. *The Fokker-Planck equation: methods of solution and applications.* Springer-Verlag, 1996.