Appendix B: System size expansion of the master equation for demographic noise

Carl Boettiger[a]

[a] *Dept of Environmental Science, Policy, and Management, University of California Berkeley, Berkeley CA 94720-3114, USA*

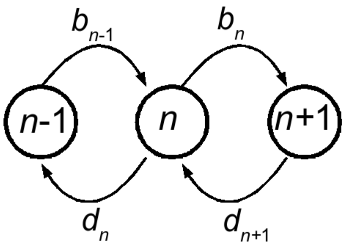

Figure 1: Birth and death as state-dependent transition rates in a Markov process.

Consider a population that is of size n with probability p_n, with birth and death rates given by b_n and d_n respectively, diagrammed in Fig~1 This representation is known as a single step Markov process. The rate of transitions to the $n+1$ state b_n, the rate of transitions to $n-1$ is d_n. Both of these are transitions away from the state p_n, hence the decrease the probability of p_n. The probability p_n increases by transitions into the state n, from either side: births enter from the state below at rate b_{n-1} and deaths from the state above, d_{n+1}. Hence the rate of change in p_n is given by

$$\dot{p}_n = b_{n-1}p_{n-1} + p_{n+1}d_{n+1} - [b_n + d_n]p_n \tag{1}$$

This probability balance is known for historical reasons as a master equation. This equation will form the center of our treatment. Master equations of this form can be written more concisely by introducing the step operator, $\mathbb{E}^k$ such that $\mathbb{E}^k f_n = f_{n+k}$, giving us

$$\frac{\mathrm{d}p_n}{\mathrm{d}t} = (\mathbb{E}^{-1} - 1)b_n p_n + (\mathbb{E} - 1)d_n p_n \tag{2}$$

which for historical reasons is known as a master equation. From this we can directly calculate the mean and variance for this process by multiplying by n or n^2 and summing over all n. These calculations are greatly simplified by observing the following property of the step operator: for any pair of test functions f_n, g_n,

$$\sum_{n=0}^{N-1} g_n \mathbb{E} f_n = \sum_{n=1}^{N} f_n \mathbb{E}^{-1} g_n$$

For example, the mean is:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\langle n\rangle &= \sum n(\mathbb{E}^{-1} - 1)b_n p_n + \sum n(\mathbb{E} - 1)d_n p_n \\ &= \sum b_n p_n(\mathbb{E} - 1)n + \sum d_n p_n(\mathbb{E}^{-1} - 1)n \\ &= -\langle d_n\rangle + \langle b_n\rangle \end{aligned} \tag{3}$$

and similarly the second moment is

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle n^2\rangle = 2\langle n(b_n - d_n)\rangle + \langle b_n + d_n\rangle \tag{4}$$

The key observation here is that the variance includes the sum of birth and death, not just the difference. Intuitively we would expect a population with large numbers of births and deaths that average to a small difference to fluctuate much more than one in which this difference is very small. Despite this simple observation, countless models purporting to include demographic noise ignore this distinction. While all moments can be derived as above, it is often impossible to solve these equations for nonlinear models. Instead, we introduce an approximation of Eq (2) that is physically motivated and intuitive. The method is based on derivation presented by the van Kampen (Kampen 2007), though a formal proof can be found in earlier work of Thomas Kurtz (Kurtz 1970; Kurtz 1971; Kurtz 1972).

The system size expansion for a birth-death process

Introduction, a change of variables

In this approximation, we expand the master equation (2) in terms of a measure of the system size, Ω. The heart of this approximation is a change of variables, the rest is simply book-keeping. We begin by explaining this change of variables, which replaces n by some average value ϕ and some fluctuations ξ.

Observe that (2) is written in terms of discrete individuals, represented by the integer n. Many population models permit real-valued variables for the population, usually interpreted as the density of individuals, for which fractional values have meaning. If we go out to the field and mark off a very large area and count all the individuals within it, we can expect to get the population density, ϕ.

Over some appropriate region, we expect the population density to be independent of our survey area. Knowing the density, we can predict how many individuals we'd expect to find in any given area Ω, simply $\phi\Omega$. We also know that the larger the area, the more accurate our prediction. We call ϕ a macroscopic variable – it describes what we expect to see over an entire population (the macroscopic level) on average, rather than at the individual level. It is an intensive (bulk) variable, because it does not depend on the area surveyed, while number n will depend on the area Ω considered. We expect n to deviate around an average value of $\phi\Omega$ by some amount that depends on the system size. For several reasons (such as the error term we found in the simple Poisson process), we will guess that the size of the fluctuations ξ scale with system size as $\Omega^{1/2}$. Mathematically,

$$n = \Omega\phi(t) + \Omega^{1/2}\xi \tag{5}$$

We will change Eq (2) into the variables ϕ and ξ. We begin with the step operator, which can be approximated by a Taylor expansion.To formulate a Taylor expansion of the step operator (where k can be a positive or negative integer), first consider the steop operator under the original discrete variable n:

Take as the definition

$$\mathbb{E}^k f(n) = f(n+k)$$

Then

$$\begin{aligned}\mathbb{E}^k n &= n + k \\ \mathbb{E}^k n^2 &= \mathbb{E}^k nn = (n+k)(n+k) = n^2 + 2kn + k^2\end{aligned}$$

This suggests we can approximate the step operator by a Taylor series

$$\mathbb{E}^k = 1 + k\frac{\partial}{\partial n} + \frac{k^2}{2}\frac{\partial^2}{\partial n^2} + \dots \tag{6}$$

To change variables, recall the chain rule,

$$\frac{\partial}{\partial \xi} f(n(\xi)) = \frac{\partial n}{\partial \xi}\frac{\partial}{\partial n} f(n(\xi))$$

hence

$$\frac{\partial}{\partial n} = \left(\frac{\partial n}{\partial \xi}\right)^{-1}\frac{\partial}{\partial \xi}$$

Making this substitution to (6) we find:

$$\mathbb{E}^k = 1 + \Omega^{-1/2}k\frac{\partial}{\partial \xi} + \Omega^{-1}\frac{k^2}{2}\frac{\partial^2}{\partial \xi^2} + \dots \tag{7}$$

Taking $P(n,t) = \Pi(\xi,t)$, we can rewrite the time derivative. First, note that the derivative of the probability distribution in the master equation, $\frac{\partial}{\partial t}P(n,t)$ is taken with n held constant,

$$\frac{\mathrm{d}n}{\mathrm{d}t} = \Omega\frac{\mathrm{d}\phi(t)}{\mathrm{d}t} + \Omega^{1/2}\frac{\mathrm{d}\xi}{\mathrm{d}t} = 0,$$

therefore

$$\frac{\mathrm{d}\xi}{\mathrm{d}t} = -\Omega^{1/2}\frac{\mathrm{d}\phi(t)}{\mathrm{d}t}. \tag{8}$$

Also by the chain rule, we have

$$\frac{\partial}{\partial t}P(n,t) = \frac{\partial \Pi}{\partial t} + \frac{\partial \Pi}{\partial \xi}\frac{\mathrm{d}\xi}{\mathrm{d}t}$$

Substituting in (8), we find:

$$\frac{\partial}{\partial t}P(n,t) = \frac{\partial \Pi}{\partial t} - \Omega^{1/2}\frac{\mathrm{d}\phi}{\mathrm{d}t}\frac{\partial \Pi}{\partial \xi} \tag{9}$$

To complete the transformation, we put (9) on the left side, replace all the $\mathbb{E}$'s in (2) with (7), and all the n's appearing in the functions b_n and d_n with (5):

$$\begin{aligned}&\frac{\partial \Pi(\xi,t)}{\partial t} - \Omega^{1/2}\frac{\mathrm{d}\phi}{\mathrm{d}t}\frac{\partial \Pi}{\partial \xi} = \\ &\quad\left(-\Omega^{-1/2}\frac{\partial}{\partial \xi} + \frac{\Omega^{-1}}{2}\frac{\partial^2}{\partial \xi^2} + \dots\right) b(\phi\Omega + \xi\Omega^{1/2})\Pi(\xi,t) \\ &+ \quad\left(\Omega^{-1/2}\frac{\partial}{\partial \xi} + \frac{\Omega^{-1}}{2}\frac{\partial^2}{\partial \xi^2} + \dots\right) d(\phi\Omega + \xi\Omega^{1/2})\Pi(\xi,t)\end{aligned} \tag{10}$$

Collecting terms of order $\Omega^{1/2}$ on both sides, we have:

$$\frac{\mathrm{d}\phi(t)}{\mathrm{d}t} = b(\phi) - d(\phi) = \alpha_1(\phi) \tag{11}$$

Where we have the part of the birth and death functions that depend on the macroscopic variable ϕ alone. This is known as the macroscopic equation, and corresponds to the density equations commonly written down. Following Kampen (2007), we will call this difference $\alpha_1(\phi)$ as a shorthand[1]

Collecting terms of order Ω^0 we recover the diffusion equation:

[1] This notation suggests it is the first jump moment, defined as first moment of the transition rate between states in a Markov process. It so happens this term will give the macroscopic law for any Markov process, not only a birth-death process. The second moment of the transition rates, a_2, will be for us simply the sum of the birth and death rates. We use this notation as it generalizes to processes that have a different master equation from (2), to describe a transition of arbitrary step size: $\alpha_{i,j} = \int r^i \Phi_j(r)\mathrm{d}r$.

$$\frac{\partial \Pi}{\partial t} = -\alpha'_{1,0}(\phi)\frac{\partial}{\partial \xi}\xi\Pi + \tfrac{1}{2}\alpha_{2,0}(\phi)\frac{\partial^2}{\partial \xi^2}\Pi \tag{12}$$

Where we define $\alpha_2(\phi) = b(\phi) + d(\phi)$ as an analogously. From this it is a straight forward exercise to calculate the moments of the distribution (multiply by ξ, ξ^2 and integrate),

$$\partial_t\langle\xi\rangle = \alpha'_{1,0}(\phi)\langle\xi\rangle \tag{13}$$
$$\partial_t\langle\xi^2\rangle = 2\alpha'_{1,0}(\phi)\langle\xi\rangle + \alpha_{2,0}(\phi) \tag{14}$$

The variance $\sigma_\xi^2 = \langle\xi^2\rangle - \langle\xi\rangle^2$ obeys the same relation as the second moment. These equations must be solved using $\phi(t)$ from the macroscopic solution, solved with the appropriate initial condition n_0. Then we can transform back to the original variables:

$$\langle n\rangle = \phi(t|n_0)\Omega + \Omega^{1/2}\langle\xi\rangle \tag{15}$$
$$\sigma^2 = \Omega\sigma_\xi^2 \tag{16}$$

It is possible to prove that any solution to (12) must be Gaussian (Kurtz 1970; Kurtz 1971). Consequently, knowing these two moments completely determines the distribution of n. This is often assumed for demographic noise. We have justified this common Gaussian noise assumption by showing it is simply a consequence of expanding the master equation to linear order, $\mathcal{O}(\Omega^0)$. From (15) we can also conclude that the macroscopic (average) variable obeys the deterministic law.

Stochastic Differential Equations

The partial differential equation (PDE) given by (12) is sometimes referred to as the Fokker-Planck equation. It is possible to prove this PDE is equivalent to as the stochastic differntial equation (SDE),

$$\mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t \tag{17}$$

which, despite the name, it is important to remember that this is actually an integral equation in the Itô calculus, and not a derivative. Note the separation of the differential elements, which are meant as shorthand for implicit integrals:

$$X_{t+s} - Xt = \int_t^{t+s} \mu(X_\tau, \tau)\mathrm{d}\tau + \int_t^{t+s} \sigma(X_\tau, \tau)\mathrm{d}B\tau \tag{18}$$

The first integral is the ordinary Lebesgue integral, while the second is an Itô integral, which integrates over the stochastic process B_t (Brownian motion, also termed a Weiner process). The Itô calculus follows different rules from the classic calculus, largely due to the fact that an Itô integral has non-zero quadratic variation. (Crudely put, $\mathrm{d}x^2 \sim 0$, but $\mathrm{d}B^2 \sim dx$). Though it is common to formulate stochastic models directly as SDEs, it is worth remembering that the SDE is the result of a small noise approximation in the system size expansion, and can break down under larger deviations, even when the mean-field equation remains approximately accurate. Ovaskainen and Meerson (2010) provides an overview literature in that addresses this limit and discusses the use of the WKB[2] method to address this case.

Fluctuation dissipation theorem

Eq (14) is particularly instructive. Note that $\alpha_2(\phi)$ is always positive, and hence will try to increase the fluctuations $\langle\xi^2\rangle$. As this term increases, it increases the influence of its coefficient $\alpha'_1(\phi)$. If this term is negative (as it must be near a stable equilibrium) then it serves to dissipate these fluctuations. Note the dissipation comes from the macroscopic behavior alone, and does not depend on knowing b and d separately. Since this dissipation is strongest for large fluctuations and small for very small fluctuations, ξ^2 will exponentially approach an equilibrium:

$$\langle\xi^2\rangle = \frac{-\alpha_2(\phi)}{2\alpha'_1(\phi)} \tag{19}$$

However, $\alpha'_1(\phi)$ need not be negative everywhere, but will be negative for any stable point, $\alpha'_1(\phi^s) < 0$. Hence there will be a region around any stable point where this fluctuation-dissipation relationship given by (19) provides a good description of the fluctuations. Consequently, for any birth-death process $b > d$ we can write down the equilibrium fluctuations as:

$$\sigma^2 = \frac{b+d}{2[d' - b']} \tag{20}$$

This expression has the wonderful properties of being both simple and general.

Example: The Levins patch model

The Levins meta-population model is given by

$$\frac{\mathrm{d}n}{\mathrm{d}t} = cn\left(1 - \frac{n}{N}\right) - en \tag{21}$$

where n is the number of occupied patches, N the total number of patches, c is the colonization rate, and e the extinction rate. The colonization term provides our birth rate function and the extinction rate the death rate function. The total number of patches N is an obvious choice for the system size Ω. From this we can immediately

[2] Wentzel, Kramers & Brillouin. Naturally this is KWB in Holland, BWK in France, and JWKB in England (for Jeffreys). Note that the physicist Gregor Wentzel is not the same as the mathematician Alexander Wentzell of the Fredilin-Wentzell theorem for large deviations years later, though I believe the former's result can be derived directly from the latter's.

apply the above theory. The jump moments written in the macroscopic variable are:

$$\alpha_1(\phi) = c\phi(1-\phi) - e\phi \tag{22}$$
$$\alpha_2(\phi) = c\phi(1-\phi) + e\phi \tag{23}$$

From which the macroscopic equation (11), the fluctuations (14) can be solved. According to (12) the distribution is simply the Gaussian with the mean given by (15) and variance (16). The equilibrium of the macroscopic equation is:

$$\langle n \rangle_s = N\left(1 - \frac{e}{c}\right)$$

while the steady-state fluctuations are given by (19), (16):

$$\sigma_n^2 = N\frac{e}{c}$$

Comparison to other models:

Note that the macroscopic equation for Levins' patch model has the same mathematical formulation as the familiar logistic equation,

$$\frac{\mathrm{d}n}{\mathrm{d}t} = rn\left(1 - \frac{n}{K}\right)$$

though the partition into birth and death rates is not explicit. Different ways of dividing this equation between birth and death can therefore create different fluctuation patterns, even as the macroscopic average remains unchanged. For instance, taking $b = rn$ and $d = rn^2/K$ and we can calculate the fluctuations of the logistic equation around its equilibrium $n^* = K$ as:

$$\sigma^2 = \frac{rn^* + rn^{*2}/K}{2(2rn^*/K - r)} = K$$

Which agrees with calculations elsewhere (Nisbet and Gurney 1982).

An alternate derivation: the diffusion approximation

A somewhat different approach to deriving a diffusion equation predominates in the literature. This approach does not make system size explicit, but instead hinges on taking a simultaneous limit of both short time and small step size. I prefer the van Kampen derivation, since it appeals to system size rather than an arbitrary small parameter going to zero. It also provides a more natural treatment for nonlinear systems, for which the diffusion equation requires extra care. With these notes in mind, the diffusion derivation usually goes as follows:

Consider small steps $x = \epsilon n$, and hence $p_n = \epsilon P_x$, where $\epsilon \ll 1$. To keep the process from slowing down we consequently have to scale up the rates by ϵ as well:

$$b_n - d_n = \frac{A_x}{\epsilon}$$

Where A is independent of ϵ. Similarly we have to scale the sum, which being proportional to the variance must be scaled by

$$b_n + d_n = \frac{B_x}{\epsilon^2}$$

We Taylor expand the step operator $\mathbb{E}$ in the new variable:

$$\mathbb{E}^k = 1 + \epsilon k \frac{\partial}{\partial x} + \frac{\epsilon^2 k^2}{2}\frac{\partial^2}{\partial x^2} + \dots$$

Then (2) becomes:

$$\begin{aligned}\frac{\partial P(x,t)}{\partial t} &= \left(\epsilon\frac{\partial}{\partial x} + \epsilon^2\tfrac{1}{2}\frac{\partial^2}{\partial x^2} + \dots\right)\left(\frac{B_x}{2\epsilon^2} - \frac{A_x}{2\epsilon}\right)P \\ &\quad + \left(-\epsilon\frac{\partial}{\partial x} + \epsilon^2\tfrac{1}{2}\frac{\partial^2}{\partial x^2} - \dots\right)\left(\frac{B_x}{2\epsilon^2} + \frac{A_x}{2\epsilon}\right)P \end{aligned} \tag{24}$$

Collecting terms of common order ϵ,

$$= -\frac{\partial A_x P}{\partial x} + \tfrac{1}{2}\frac{\partial^2 B_x P}{\partial x^2} + \mathcal{O}(\epsilon^2) \tag{25}$$

For small ϵ we can ignore the terms of $\mathcal{O}(\epsilon^2)$, and we have recovered the diffusion approximation. Unfortunately, nature does not give us a parameter ϵ that goes to zero. We expect that larger populations will have relatively smaller fluctuations (recall the introductory example) and hence this limit should have something to do with increasing the system size. We make these notions more precise by expanding equation (2) explicitly in terms of the system size. This will lead us to a much richer and more satisfying description of demographic noise than the mathematical limit above.

System size expansion in higher-dimensional systems: environmental noise example

We introduce environmental noise into the master equation by allowing the transition rates to depend explicitly on the macroscopic variable ψ as well as the state variable ϕ. It can be helpful to think of fluctuations in ψ as also arising from some lower level transitions, but we need not model these explicitly. Following the expansion as before (see Kampen (2007) for details), the mean dynamics are given by the macroscopic equations:

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = \alpha_{1,0}(\phi, \psi)$$
$$\frac{\mathrm{d}\psi}{\mathrm{d}t} = \beta_{1,0}(\psi)$$

while the fluctuation dynamics are now given by three coupled differential equations:

$$\frac{\mathrm{d}\langle \xi^2 \rangle}{\mathrm{d}t} = 2\frac{\partial \alpha_{1,0}}{\partial \phi}\langle \xi^2 \rangle + 2\frac{\partial \alpha_{1,0}}{\partial \psi}\langle \xi\eta \rangle + \alpha_{2,0} \tag{26}$$
$$\frac{\mathrm{d}\langle \xi\eta \rangle}{\mathrm{d}t} = \left(\frac{\partial \alpha_{1,0}}{\partial \phi} + \frac{\partial \beta_{1,0}}{\partial \psi}\right)\langle \xi\eta \rangle + \frac{\partial \alpha_{1,0}}{\partial \psi}\langle \eta^2 \rangle \tag{27}$$
$$\frac{\mathrm{d}\langle \eta^2 \rangle}{\mathrm{d}t} = 2\frac{\partial \beta_{1,0}}{\partial \psi}\langle \eta^2 \rangle + \beta_{2,0} \tag{28}$$

Without an explicit model for the environmental variation process, we can merely summarize its role in terms of the autocorrelation, $\tau_c := \frac{\partial \beta_{1,0}}{\partial \psi}$ and its overall variance at steady-state, $\langle \eta^2 \rangle = \frac{\beta_{2,0}}{-2\partial_\psi \beta 1,0}$, which we will refer to as the environmental variation $\sigma_e^2 := \langle \eta^2 \rangle$. A more detailed mechanistic description of how the environmental noise arises would permit a time-dependent solution, but the system size expansion tells us that these two summary statistics are all we need to express the steady-state solution above.

Setting the remaining equations to zero to solve for the stationary state we can write down a general formula for the variation in ϕ:

$$\langle xi^2 \rangle = \frac{\left(\partial_\psi \alpha_{1,0}(\phi, \psi)\right)^2}{\left(\partial_\phi \alpha_{1,0}(\phi, \psi)\right)^2 + \partial_\phi \alpha_{1,0}(x, y)\tau_c}\sigma_e^2 + \sigma_d^2$$

From which we recover the environmental variation equation presented in the main text.

References

Kampen, N.G. van. 2007. *Stochastic Processes in Physics and Chemistry*. Third. Amsterdam, New York, Oxford: North-Holland Publishing Company.

Kurtz, T. G. 1970. "Solutions of Ordinary Differential Equations as Limits of Pure Jump Markov Processes." *Journal of Applied Probability* 7 (1): 49–58.

———. 1971. "Limit Theorems for Sequences of Jump Markov Processes Approximating Ordinary Differential Processes." *Journal of Applied Probability* 8 (2): 344–56.

———. 1972. "The Relationship between stochastic and Deterministic Models for Chemical Reactions." *The Journal of Chemical Physics* 57 (7): 2976–8.

Nisbet, R. M., and W.C.S. Gurney. 1982. *Modelling Fluctuating Populations*. New Jersey: The Blackburn Press.

Ovaskainen, Otso, and Baruch Meerson. 2010. "Stochastic models of population extinction." *Trends in Ecology & Evolution* 25 (11). Elsevier Ltd: 643–52. doi:10.1016/j.tree.2010.07.009.