9. Compositional processes

So far in this course, we have seen compositions as samples drawn from an underlying distribution. However, just like real vectors, compositions can be parameterized, either by a continuous or a discrete parameter, typically representing time or space. One such example is the data set with the world's energy production between 1960 and 2020 used in the previous chapter.

In some cases, the process is known and can be used to predict or model the evolution of a system, while in other cases, the process is unknown, but obtainable from existing compositional data.

9.1 Time-dependent compositions

A D-part composition $\mathbf{z}$ that evolves in time t is said to be time-dependent and it can be described as

$$\mathbf{x}(t) = \mathscr{C}[\mathbf{z}(t)], \qquad z_i \in \mathbb{R}_+^D \tag{9.1}$$

The dependency can be arbitrarily complex, but among the simplest cases is proportional growth (or decay), where the change in some property z depends proportionally on z itself,

$$\frac{dz}{dt} = \lambda z \tag{9.2}$$

where the sign of the rate parameter λ determines whether we have growth or decay and the magnitude of λ determines how fast it goes. Well-known examples of such processes are radioactive decay and bacterial growth. The solution to Eq. 9.2 is an exponential function,

$$z(t) = z_0 \cdot \exp(\lambda t), \tag{9.3}$$

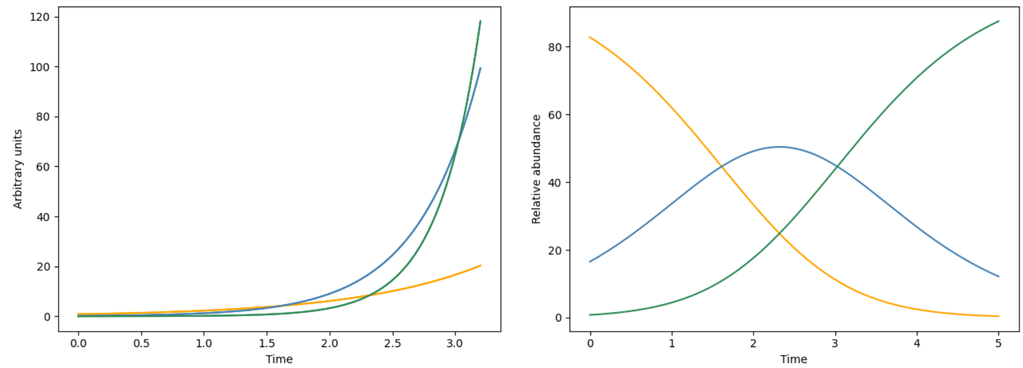

Figure 9.1: The growth in time of three different species of bacteria. The left panel shows the absolute abundances, while the right panel shows the relative abundances.

where z_0 is known as the initial condition or "starting amount". In the case of proportional growth, growth happens exponentially.

If we consider radioactive decay and we have a medium containing multiple radioactive isotopes, we can describe the system by letting z and λ in Eq. 9.2 be vectors with as many entries as there are different isotopes. We can then measure the amount of each isotope z_i, for instance using a mass spectrometer, at $t = 0$ and again at some later time $t > 0$ and solve for each λ_i to obtain their decay rates.

But we can also choose to take a compositional approach to the problem by determining relative abundances rather than absolute abundances. In this case, $\mathbf{z}$ is a composition, which may be closed to 100 for convenience. When $\mathbf{z}$ is a composition, the solution Eq. 9.3 is no longer valid. Recall from Chapter 3 that multiplication (and addition) are not valid operations on the simplex, so we must substitute the operations with their Aichison equivalents (Def. 3.2.1). Doing so gives us the compositional solution,

$$\mathbf{x}(t) = \mathbf{x}_0 \oplus t \odot \mathbf{p}, \qquad \mathbf{p} = \exp(\lambda). \tag{9.4}$$

This solution is recognized as a straight line (in the simplex), with $\mathbf{x}_0$ as the intercept, $\mathbf{p}$ as the slope, and t as the variable. From this, we can go to coordinate space using the ILR transformation, defining $\mathbf{u} = \text{ilr}(\mathbf{x})$ and $\mathbf{v} = \text{ilr}(\mathbf{p})$, so that,

$$\mathbf{u}(t) = \mathbf{u}_0 + t \cdot \mathbf{v}, \tag{9.5}$$

a straight line in coordinate space $\mathbb{R}^{D-1}$.

■ **Example 9.1 — Bacterial growth.** Three species of bacteria (x_1, x_2, x_3) grow at rates $\lambda = (1, 2, 3)$. The initial relative abundances of the three species are $\mathbf{z}_0 = (82.7\%, 16.5\%, 0.8\%)$. We can easily graph the growth of each species using the solution Eq. 9.3, which is shown in the left panel of Fig. 9.1. The right panel in that figure shows the relative abundance in time, where the solution has been closed to 100 at each point in time.

The growth process can also be shown in a ternary diagram using the compositional solution Eq. 9.4. In the ternary diagram we can track the evolution of the composition along the line. This is shown in Fig. 9.2, where the starting point $t = 0$ and the end point $t = 5$ have been labeled. Equivalently, we can plot this in a coordinate plot, like the ones

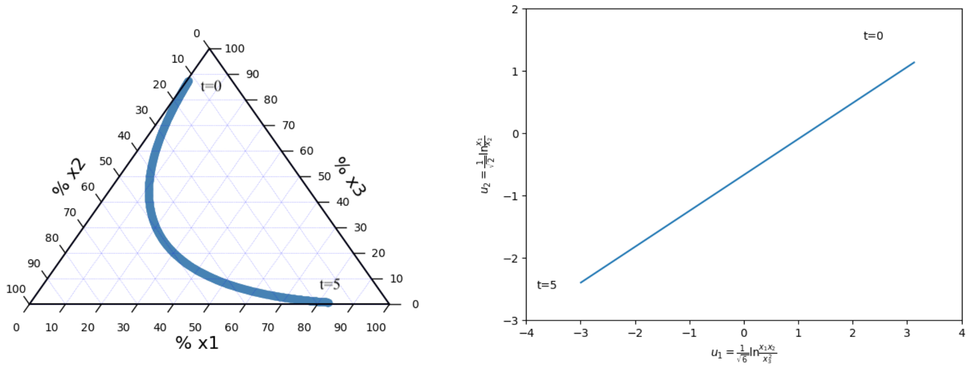

Figure 9.2: The growth in time of three different species of bacteria in a ternary diagram (left) and in coordinate space (right).

we introduced in Sect. 5.3.2. In order to obtain that plot, we need a basis, which we have chosen to be the normalized partition $((+1,+1,-1),(+1,-1,0))$. From the coordinate plot, it is very clear that we are dealing with a linear compositional process.

■

9.2 Compositional derivatives

An important part of the field of calculus, and in particular differential equations, is the derivative, which describes the rate of change of a function with respect to a variable. The derivative is defined as the difference of two slightly offset function values in the limit where the offset approaches zero. This is written as

$$\frac{df(t)}{dt} = \lim_{h \to 0} \frac{1}{h}\left(f(t+h) - f(t)\right) \tag{9.6}$$

This usual form is defined for a real function f and the derivative of f will likewise take real values. If we want to define derivatives for compositions, we need to recall that the sample space for compositions is the simplex, and thus we have to use the homologous algebraic operations, which we introduced in chapter 3. We can see that the expression above contains the inverse sum (minus) of two function values as well as an inverse scale multiplication (divide by h). Defining derivatives for composition thus comes down to replacing those operations by (inverse) perturbation and (inverse) powering,

$$\frac{d_{\oplus}\mathbf{x}(t)}{dt} = \lim_{h \to 0} \frac{1}{h} \odot \left(\mathbf{x}(t+h) \ominus \mathbf{x}(t)\right), \tag{9.7}$$

where $\mathbf{x}$ is a compositional process. We use the subscript $\oplus$ to denote the compositional derivative. However, this definition is not practical for calculating actual derivatives, but we can apply our usual strategy of transforming the compositions into coordinate space, carry out regular real calculus, and back-transform into the simplex,

$$\frac{d_{\oplus}\mathbf{x}(t)}{dt} = ilr^{-1}\left[\frac{d\, ilr(\mathbf{x}(t))}{dt}\right] \tag{9.8}$$

The CLR transformation can be used likewise, since CLR is also isometric. However, we can not use ALR for doing derivatives because it is not isometric. The CLR version is similar to the ILR version, but due to the definition of CLR, the expression can be simplified to,

$$\frac{d_{\oplus}\mathbf{x}(t)}{dt} = clr^{-1}\left[\frac{d\ clr(\mathbf{x}(t))}{dt}\right] = \mathscr{C}\left[\exp\left(\frac{d\ \ln(\mathbf{x}(t))}{dt}\right)\right]. \tag{9.9}$$

The compositional derivative has analogue qualities to what we are used to from real calculus. It is well known that if we differentiate a constant using the normal derivative, we get zero. Likewise, if we take the compositional derivative of a constant composition, i.e., a composition that is not part of a process, we get $\mathbf{n} = (1/D, 1/D, ...)$, the center of the simplex. Recall from Sect. 5.3.1 how the center of the simplex acts as a neutral element like 0 does in real algebra.

The above result is a special case of the well-known rule for differentiating polynomials. The standard rule in normal calculus is,

$$\frac{dP_m(x)}{dx} = \sum_{k=1}^{m} k a_k x^{k-1} \tag{9.10}$$

for an m-degree polynomial with coefficient vector $\mathbf{a} = (a_1, a_2, ..., a_m)$. The compositional version applies when the vector $\mathbf{a}$ is a composition, in which case, after replacing the operators by the compositional homologous, we get,

$$\frac{d_{\oplus}P_m(x)}{dx} = \bigoplus_{k=1}^{m} k x^{k-1} \odot \mathbf{a}. \tag{9.11}$$

9.3 Compositional differential equations

Above, in Sect. 9.1, we encountered the simplest possible differential equation. The general form of these first-order differential equations is

$$\frac{d_{\oplus}\mathbf{x}(t)}{dt} = \mathbf{f}(\mathbf{x}(t), t), \tag{9.12}$$

that is, the compositional derivative of a compositional process equals a simplex-valued function of the process itself and, in general, time. When $\mathbf{f}$ is not explicitly dependent on time t, the equation describes an isolated system, free of external forcing. Examples of external forcing would be a damped oscillator or, in the case of bacterial growth, the influence of anti-microbials.

In general, we can solve Eq. 9.12 by ILR transforming both sides of the equation. Using Eq. 9.8 and denoting the ILR-corrdinates by $\mathbf{x}^*$, we can write Eq. 9.12 as,

$$\frac{d\mathbf{x}^*(t)}{dt} = \mathbf{f}^*(\mathbf{x}^*(t), t). \tag{9.13}$$

This equation describes a system of $D-1$ ordinary differential equations, which can be solved using standard techniques. Equation 9.13 will in general have a different expression depending on the ILR basis used for the transformation, but every solution to Eq. 9.13 will inversely ILR transform back to the same process $\mathbf{x}(t)$.

When the right-hand side of the differential equation is a constant composition, $\mathbf{z}$, with $z_i = exp(\lambda_i)$, we saw above that the solution is a straight line in the simplex, in which case the differential equation can be written as

$$\begin{aligned} \frac{d_\oplus \mathbf{x}(t)}{dt} &= \mathscr{C}exp(\lambda) && (9.14)\\ \mathbf{x}(t) &= \mathbf{a} \oplus t \odot \mathscr{C}exp(\lambda) = \mathbf{a} \oplus \mathscr{C}exp[\lambda_1 t, \lambda_2 t, ..., \lambda_n t]. && (9.15) \end{aligned}$$

The following example is an application of this equation.

9.3.1 Population dynamics

Early in the 19th century, a discussion was raised between two mathematicians, the Englishman Thomas Malthus and the Belgian Pierre Verhulst, about the future of the human population. Given the exponential increase in population at the time, Malthus proposed that the human population would continue to grow exponentially. Obviously, the Earth is not capable of sustaining an exponentially growing population forever, so Verhulst counter proposed that the population growth would follow the logistic curve, that is, flatten out at a certain constant population. The controversy at the time centered around which differential equation would rightly describe the growth of the human population. Malthus argued that population growth was governed by the equation,

$$\frac{dN}{dt} = N, \qquad N = N_0 exp(t), \tag{9.16}$$

whereas Verhulst proposed the logistic equation and its solution,

$$\frac{dN}{dt} = \alpha N - \beta N^2, \qquad N = \frac{K_1}{1 + K_2 exp(-\alpha t)}, \tag{9.17}$$

as the best way to describe population growth.

As it turns out, when considering the problem from a compositional perspective, both scenarios are solutions to the same compositional differential equation, and the only difference between the solutions is the assumption on the availability of resources. The total amount of available resources M at a given time t can be described as the sum of consumed C and remaining R resources,

$$M(t) = C(t) + R(t), \tag{9.18}$$

and the idea is that consumed resourced is a proxy for the number of people alive. Because consumed and remaining resources sum up to a constant, they can be regarded as parts of a two-part composition. The resource compositional process can thus be written as,

$$\mathbf{x}(t) = M(t)[C(t), R(t)]. \tag{9.19}$$

If we consider the simplicial differential equation and its solution, Eq. 9.14,

$$\begin{aligned} \frac{d_\oplus \mathbf{x}(t)}{dt} &= \mathscr{C}[exp(\lambda_1), exp(\lambda_2)] && (9.20)\\ \mathbf{x}(t) &= \mathbf{a} \oplus \mathscr{C}[exp(\lambda_1 t), exp(\lambda_2 t)] && (9.21)\\ &= \left[\frac{a_1 exp(\lambda_1 t)}{a_1 exp(\lambda_1 t) + a_2 exp(\lambda_2 t)}, \frac{a_2 exp(\lambda_2 t)}{a_1 exp(\lambda_1 t) + a_2 exp(\lambda_2 t)}\right] && (9.22) \end{aligned}$$

we can identify both the Malthus and the Verhulst solutions by a proper choice of λ_2 and $M(t)$. If we let $\lambda_2 = 0$ and substitute the solution into eq. 9.19, we get,

$$\mathbf{x}(t) = M(t)\left[\frac{a_1 exp(\lambda_1 t)}{a_1 exp(\lambda_1 t) + a_2}, \frac{a_2}{a_1 exp(\lambda_1 t) + a_2}\right] \tag{9.23}$$

By letting $M(t) = (a_1 exp(\lambda_1 t) + a_2)/a_2$, the solution reduces to,

$$\mathbf{x}_{Malthus}(t) = [a_1/a_2 exp(\lambda_1 t), 1], \tag{9.24}$$

which is exactly exponential growth in the consumed resources (the population), while the remaining resources stay constant. On the other hand, we can also just let $M(t)$ be a constant (equal to 1 for simplicity), in which case, after dividing the first part by $exp(\lambda_1 t)$, the solution becomes,

$$\mathbf{x}_{Verhulst}(t) = \left[\frac{a_1}{a_1 + a_2 exp(-\lambda_1 t)}, \frac{a_2}{a_1 exp(\lambda_1 t) + a_2}\right] \tag{9.25}$$

The first part of this solution, which describes the population growth, is clearly recognized as having the same algebraic form as the solution in Eq. 9.17, the Verhulst solution.

So both scenarios can be derived from the same underlying compositional process. The difference between them is that Malthus assumed that there would always be a constant amount of remaining resources or an infinite amount of available resources, while Verhulst assumed that the sum of remaining and consumed resources would be constant, corresponding to a finite amount of available resources.

9.3.2 Epidemics

One of the standard tools used by epidemiologists when facing a new epidemic is the compartment model, also known as SIR, where S, I, and R stand for susceptible, infected, and recovered. SIR is the simplest compartment model, and very often, additional compartments are introduced, such as exposed, contagious, susceptible-again, and so on. Here we shall consider the SIR model for an epidemic, where people are contagious while they are infected and recovered people cannot be infected again.

The basic form of the model is a differential equation, where the sum of changes in the compartment equals 0,

$$\frac{dS(t)}{dt} + \frac{dI(t)}{dt} + \frac{dR(t)}{dt} = 0. \tag{9.26}$$

Since the number of people is assumed to be constant (no deaths or births), the solution to this equation must have the form,

$$S(t) + I(t) + R(t) = N, \tag{9.27}$$

where N is the size of the population. Equation 9.27 shows that the three compartments S, I, and R form a composition (because they sum up to a constant), and because they are all functions of time, the model is a compositional process.

The three derivatives in eq. 9.26 depends on two rates: the rate of infection or contagiousness β and the rate of recovery γ, with the basic reproduction number defined as

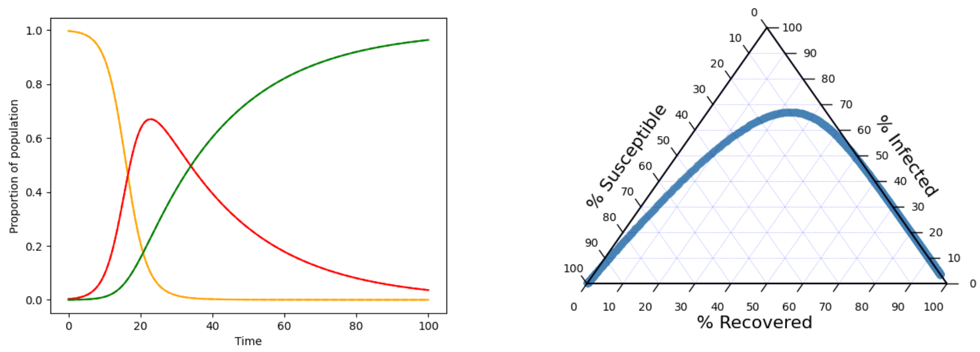

Figure 9.3: The evolution of an epidemic according to the SIR model. Orange shows the susceptible, red are the infected, and green are the recovered population. The right panel shows the process in the simplex.

$R_0 = \beta/\gamma$, the number of new infections caused by one infected individual in a susceptible population. The first rate, β, determines how fast people move from S to I, while the second rate, γ, determines how fast they move from I to R. In the compositional formulation, the three derivatives can be expressed as

$$\frac{dS(t)}{dt} = -\beta S(t)I(t) \qquad \frac{dI(t)}{dt} = \beta S(t)I(t) - \gamma I \qquad \frac{dR(t)}{dt} = \gamma I(t), \tag{9.28}$$

where S, I, and R are the proportion of people in each compartment at any given time t. This is an example of a system of non-linear, first-order, compositional differential equations, that cannot be solved analytically, except for a few trivial cases, such as, when the number of infected is zero and everything stays constant. In this case, the system of equations is also no longer compositional. The non-linearity comes from the mixed term in the first two equations, which involves the product $S \times I$.

Because of the compositional nature of the process, we only need to solve two equations because the third part is determined from the other two through closure, e.g., $R(t) = \kappa - [S(t) + I(t)]$, where κ is the closure constant. Because of this, we can obtain a so-called implicit solution by dividing the equation for dS/dt by the equation for dR/dt,

$$\frac{dS(t)}{dR(t)} = -R_0 S(t), \tag{9.29}$$

which have the implicit solution,

$$S(t) = S(0)e^{-R_0(R(t)-R(0))}. \tag{9.30}$$

A full solution requires numerical integration, and an example solution for a typical value of R_0 of 10, can be seen in Fig. 9.3.

9.4 Exercises

Exercise 9.1 A mineral assemblage contains three radioactive isotopes,

$$[^{238}\mathrm{U}, ^{232}\mathrm{Th}, ^{40}\mathrm{K}] = [150, 30, 120] \text{ ppm}$$

at $t = 0$. The half-lives are $4.468 \cdot 10^9, 14.05 \cdot 10^9$, and $1.277 \cdot 10^9$ years, respectively. Plot the evolution of the composition over the course of $50 \cdot 10^9$ years.
HINT: Half-life, $t_{1/2}$, is related to decay rate as $t_{1/2} = \frac{log(2)}{\lambda}$.
HINT: measure time in units of 10^9 years to avoid numeric underflow ■

Exercise 9.2 Consider the compositional differential equation,

$$\frac{d_{\oplus}\mathbf{x}(t)}{dt} = \mathbf{x} \boxdot \mathbf{A} \oplus \mathbf{f}$$

with

$$\mathbf{A} = \begin{bmatrix} 0.56 & 2.55 & -3.11 \\ -1.40 & -1.61 & 3.01 \\ 0.84 & -0.94 & 0.10 \end{bmatrix}, \qquad \mathbf{f}^T = \begin{bmatrix} 0.37 \\ 0.03 \\ 0.60 \end{bmatrix}$$

Using the sequential binary partition,

$$\Psi = \text{norm}\left(\begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 0 \end{bmatrix}\right)$$

calculate the contrast matrix,

$$\mathbf{A}^* = \Psi \mathbf{A} \Psi^T$$

and its eigenvalues. Obtain the equilibrium solution, $(\partial x = 0)$, for the real-valued equation,

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \mathbf{x}^* \mathbf{A}^* + \mathbf{f}^*$$

This point is a fixed-point and is transformed back to the simplex, it corresponds to the equilibrium composition of the process. From the eigenvalues, what can you say about the nature of the fixed point? Obtain the full solution (optional).
HINT: Look up "phase portrait behavior" in wikipedia.
■

Exercise 9.3 The SIR model in Sect. 9.3.2 does not take births and deaths into account, and the epidemic therefore dies out when everyone has been infected. We can include births and deaths and thereby replenish the susceptible compartment. For simplicity, we let the birth rate be equal to the death rate so that the population size is constant, i.e., the composition is always closed to the same number. The SIR model with demographics is given by

$$\frac{dS}{dt} = \mu - \beta SI - \mu S \qquad \frac{dI}{dt} = \beta SI - \gamma I - \mu I \qquad \frac{dR}{dt} = \gamma I - \mu R$$

Solve this system using the supplied Python script, and plot the evolution of the model.
■