Chapter 6

Sequences and Series of Functions

6.1 Discussion: Branching Processes

The fact that polynomial functions are so ubiquitous in both pure and applied analysis can be attributed to any number of reasons. They are continuous, infinitely differentiable, and defined on all of $\mathbf{R}$. They are easy to evaluate and easy to manipulate, both from the points of view of algebra (adding, multiplying, factoring) and calculus (integrating, differentiating). It should be no surprise, then, that even in the earliest stages of the development of calculus, mathematicians experimented with the idea of extending the notion of polynomials to functions that are essentially polynomials of infinite degree. Such objects are called *power series*, and are formally denoted by

$$\sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + a_3 x^4 + \cdots .$$

The basic dilemma from the point of view of analysis is deciphering when the desirable qualities of the limiting functions (the polynomials in this case) are passed on to the limit (the power series). To put the discussion in a more concrete context, let's look at a particular problem from the theory of probability.

In 1873, Francis Galton asked the London Mathematical Society to consider the problem of the survival of surnames (which at that time were passed to succeeding generations exclusively by adult male children). "Assume," Galton said, "that the law of population is such that, in each generation, p_0 percent of the adult males have no male children who reach adult life; p_1 have one such male child; p_2 percent have two; and so on... Find [the probability that] the surname will become extinct after r generations." We should add (or make explicit) the assumption that the lives of each offspring, and the descendants thereof, proceed independently of the fortunes of the rest of the family.

Galton asks for the probability of extinction after r generations, which we will call d_r. If we begin with one parent, then $d_1 = p_0$. If $p_0 = 0$, then d_r will clearly equal 0 for all generations r. To keep the problem interesting, we will insist that from here on $p_0 > 0$. Now, d_2, whatever it equals, will certainly satisfy $d_1 \leq d_2$ because if the population is extinct after one generation it will remain so after two. By this reasoning, we have a monotone sequence

$$d_1 \leq d_2 \leq d_3 \leq d_4 \cdots,$$

which, because we are dealing with probabilities, is bounded above by 1. By the Monotone Convergence Theorem, the sequence converges, and we can let

$$d = \lim_{r \to \infty} d_r$$

be the probability that the surname eventually goes extinct at any time in the future. Knowing it exists, our task is to find d.

The truly clever step in the solution is to define the function

$$G(x) = p_0 + p_1 x + p_2 x^2 + p_3 x^3 + \cdots.$$

In the case of producing male offspring, it seems safe to assume that this sum terminates after five or six terms, because nature would have it that $p_n = 0$ for all values of n beyond this point. However, if we were studying neutrons in a nuclear reactor, or heterozygotes carrying a mutant gene (as is often the case with the theory of branching processes), then the notion of an infinite sum becomes a more attractive model. The point is this: We will proceed with reckless abandon and treat the function $G(x)$ as though it were a familiar polynomial of finite degree. At the end of the computations, however, we will have to again become well-trained analysts and be prepared to justify the manipulations we have made under the hypothesis that $G(x)$ represents an infinite sum for each value of x.

The critical observation is that

$$G(d_r) = d_{r+1}.$$

The way to understand this is to view the expression

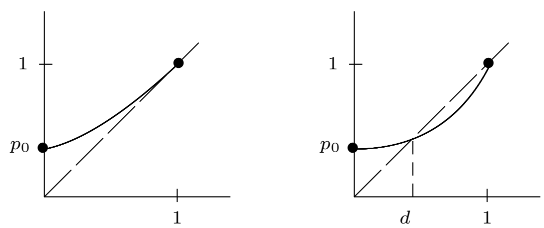
$$G(d_r) = p_0 + p_1 d_r + p_2 d_r^2 + p_3 d_r^3 + \cdots$$

as a sum of the probabilities for different distinct ways extinction could occur in $r + 1$ generations based on what happens after the first generational step. Specifically, p_0 is the probability that the initial parent has no offspring and so still has none after $r + 1$ generations. The term $p_1 d_r$ is the probability that the initial parent has one male child times the probability that this child's own lineage dies out after r generations. Thus, the probability $p_1 d_r$ is another contribution toward the probability of extinction in $r + 1$ steps. The third term represents the probability that the initial parent has two children and that the surnames of each of these two children die out within r generations. Continuing in this way, we see that every possible scenario for extinction in $r + 1$ steps is accounted for exactly once within the sum $G(d_r)$. By the definition of d_{r+1}, we get $G(d_r) = d_{r+1}$.

Now for some analysis. If we take the limit as $r \to \infty$ on each side of the equation $G(d_r) = d_{r+1}$, then on the right-hand side we get $\lim d_{r+1} = d$. Assuming G is continuous, we have

$$d = \lim_{r\to\infty} d_{r+1} = \lim_{r\to\infty} G(d_r) = G(d).$$

The conclusion that $d = G(d)$ means that the point d is a *fixed point* of G. It can be located graphically by finding where the graph of G intersects the line $y = x$.



(i) $G'(1) \leq 1$ (ii) $G'(1) > 1$

It is always the case that

$$G(1) = p_0 + p_1 + p_2 + p_3 + \cdots = 1$$

because the probabilities (p_k) form a complete distribution. But $d = 1$ is not necessarily the only candidate for a solution to $G(d) = d$. Graph (ii) illustrates a scenario in which G has another fixed point in the interval $(0, 1)$ in addition to $x = 1$.

Treating G as though it were a polynomial, we differentiate term-by-term to get

$$G'(x) = p_1 + 2p_2x + 3p_3x^2 + 4p_4x^3 + \cdots$$

and

$$G''(x) = 2p_2 + 6p_3x + 12p_4x^2 + \cdots .$$

On the interval $[0, 1]$, every term in G' and G'' is nonnegative which means G is an increasing, convex function from $G(0) = p_0 > 0$ up to $G(1) = 1$. This suggests that the two preceding graphs form a rather complete picture of the possibilities for the behavior of G with regard to fixed points. Of particular interest is graph (ii), where the graph of $y = x$ intersects G twice in $[0, 1]$. Using the Mean Value Theorem, we can prove (Exercise 5.3.6) that $G(d) = d$ for some other point $d \in (0, 1)$ if and only if $G'(1) > 1$.

Now,

$$G'(1) = p_1 + 2p_2 + 3p_3 + 4p_4 + \cdots$$

has a very interesting interpretation within the language of probability. The sum is a weighted average, where in each term we have multiplied the number of male children by the probability of actually producing this particular number. The result is a value for the *expected number of male offspring* from a given parent. Said another way, $G'(1)$ is the average number of male children produced by the parents in this particular family tree.

It is not difficult to argue that (d_r) will converge to the smallest solution to $G(d) = d$ on $[0, 1]$ (Exercise 6.5.12), and so we arrive at the following conclusion. If each parent produces, on average, more than one male child, then there is a positive probability that the surname will survive. The equation $G(d) = d$ will have a unique solution in $(0, 1)$, and $1 - d$ represents the probability that the surname does not become extinct. On the other hand, if the expected number of male offspring per parent is one or less than one, then extinction occurs with probability one.

The implications of these results on nuclear reactions and the spread of cancer are fascinating topics for another time. What is of concern to us here is whether our manipulations of $G(x)$ are justified. The assumption that $\sum_{n=0}^{\infty} p_n = 1$ guarantees that G is at least defined at $x = 1$. The point $x = 0$ poses no problem, but is G necessarily well-defined for $0 < x < 1$? If so, how might we prove that G is continuous on this set? Differentiable? Twice-differentiable? If G is differentiable, can we compute the derivative by naively differentiating each term of the series? Our initial attack on these questions will require us to focus attention on the interval $[0, 1)$. Some interesting subtleties arise when we try to extend our results to include the endpoint $x = 1$.

6.2 Uniform Convergence of a Sequence of Functions

Just as in chapter two, we will initially concern ourselves with the behavior and properties of converging *sequences* of functions. Because convergence of infinite sums is defined in terms of the associated sequence of partial sums, the results from our study of sequences will be immediately applicable to the questions we have raised about power series and about infinite series of functions in general.

Pointwise Convergence

Definition 6.2.1. For each $n \in \mathbf{N}$, let f_n be a function defined on a set $A \subseteq \mathbf{R}$. The sequence (f_n) of functions *converges pointwise on* A to a function $f : A \to \mathbf{R}$ if, for all $x \in A$, the sequence of real numbers $f_n(x)$ converges to $f(x)$.

In this case, we write $f_n \to f$, $\lim f_n = f$, or $\lim_{n\to\infty} f_n(x) = f(x)$. This last expression is helpful if there is any confusion as to whether x or n is the limiting variable.

Example 6.2.2. (i) Consider

$$f_n(x) = (x^2 + nx)/n$$