# Chapter 4

# One-Parameter Processes, Usually Functions of Time

Section 4.1 defines one-parameter processes, and their variations (discrete or continuous parameter, one- or two- sided parameter), including many examples.

Section 4.2 shows how to represent one-parameter processes in terms of "shift" operators.

We've been doing a lot of pretty abstract stuff, but the point of this is to establish a common set of tools we can use across many different concrete situations, rather than having to build very similar, specialized tools for each distinct case. Today we're going to go over some examples of the kind of situation our tools are supposed to let us handle, and begin to see how they let us do so. In particular, the two classic areas of application for stochastic processes are dynamics (systems changing over time) and inference (conclusions changing as we acquire more and more data). Both of these can be treated as "one-parameter" processes, where the parameter is time in the first case and sample size in the second.

## 4.1   One-Parameter Processes

The index set $T$ isn't, usually, an amorphous abstract set, but generally something with some kind of topological or geometrical structure. The number of (topological) dimensions of this structure is the number of *parameters* of the process.

**Definition 34 (One-Parameter Process)** *A process whose index set $T$ has one dimension is a* one-parameter process. *A process whose index set has more than one dimension is a* multi-parameter process. *A one-parameter process is*

discrete *or* continuous *depending on whether its index set is countable or uncountable. A one-parameter process where the index set has a minimal element it is* one-sided, *otherwise it is* two-sided.

$\mathbb{N}$ is a one-sided discrete index set, $\mathbb{Z}$ a two-sided discrete index set, $\mathbb{R}^+$ (including zero!) is a one-sided continuous index set, and $\mathbb{R}$ a two-sided continuous index set.

Most of this course will be concerned with one-parameter processes, which are intensely important in applications. This is because the one-dimensional parameter is usually either time (when we're doing dynamics) or sample size (when we're doing inference), or both at once. There are also some important cases where the single parameter is space.

**Example 35 (Bernoulli process)** *You all know this one: a one-sided infinite sequence of independent, identically-distributed binary variables, where, for all t, $X_t = 1$ with probability p.*

**Example 36 (Markov models)** *Markov chains are discrete-parameter stochastic processes. They may be either one-sided or two-sided. So are Markov models of order k, and hidden Markov models. Continuous-time Markov processes are, naturally enough, continuous-parameter stochastic processes, and again may be either one-sided or two-sided.*

Instances of physical processes that may be represented by Markov models include: the positions and velocities of the planets; the positions and velocities of molecules in a gas; the pressure, temperature and volume of the gas (Keizer, 1987); the position and velocity of a tracer particle in a turbulent fluid flow; the three-dimensional velocity field of a turbulent fluid (Frisch, 1995); the gene pool of an evolving population (Fisher, 1958; Haldane, 1932; Gillespie, 1998). Instances of physical processes that may be represented by hidden Markov models include: the spike trains of neurons (Tuckwell, 1989); the sonic wave-forms of human speech (Rabiner, 1989); many economic and social time-series (Durbin and Koopman, 2001); etc.

**Example 37 ("White Noise" (Not Really))** *For each $t \in \mathbb{R}^+$, let $X_t \sim \mathcal{N}(0, 1)$, all mutually independent of one another. This is a process with a one-sided continuous parameter.*

It would be character building, at this point, to convince yourself that the process just described exists. (You will need the Kolmogorov Extension Theorem, 29).

**Example 38 (Wiener Process)** *Here $T = \mathbb{R}^+$ and $\Xi = \mathbb{R}$. The* Wiener process *is the continuous-parameter random process where*

*1. $W(0) = 0$,*

2. *for any three times, $t_1 < t_2 < t_3$, $W(t_2) - W(t_1)$ and $W(t_3) - W(t_2)$ are independent (the "independent increments" property),*

3. *$W(t_2) - W(t_1) \sim \mathcal{N}(0, t_2 - t_1)$ and*

4. *$W(t, \omega)$ is a continuous function of t for almost all $\omega$.*

*We will spend a lot of time with the Wiener process, because it turns out to play a role in the theory of stochastic processes analogous to that played by the Gaussian distribution in elementary probability — the easily-manipulated, formally-nice distribution delivered by limit theorems.*

When we examine the Wiener process in more detail, we will see that it almost never has a derivative. Nonetheless, in a sense which will be made clearer when we come to stochastic calculus, the Wiener process can be regarded as the integral over time of something *very like* white noise, as described in the preceding example.

**Example 39 (Logistic Map)** *Let $T = \mathbb{N}$, $\Xi = [0, 1]$, $X(0) \sim U(0, 1)$, and $X(t+1) = aX(t)(1 - X(t))$, $a \in [0, 4]$. This is called the logistic map. Notice that all the randomness is in the initial value $X(0)$; given the initial condition, all later values $X(t)$ are fixed. Nonetheless, this is a Markov process, and we will see that, at least for certain values of a, it satisfies versions of the laws of large numbers and the central limit theorem. In fact, large classes of deterministic dynamical systems have such stochastic properties.*

**Example 40 (Symbolic Dynamics of the Logistic Map)** *Let $X(t)$ be the logistic map, as in the previous example, and let $S(t) = 0$ if $X(t) \in [0, 0.5)$ and $S(t) = 1$ if $X(t) = [0.5, 1]$. That is, we partition the state space of the logistic map, and record which cell of the partition the original process finds itself in. $X(t)$ is a Markov process, but these "symbolic" dynamics are not necessarily Markovian. We will want to know when functions of Markov processes are themselves Markov. We will also see that there is a sense in which, Markovian or not, this partition is exactly as informative as the original, continuous state — that it is generating. Finally, when $a = 4$ in the logistic map, the symbol sequence is actually a Bernoulli process, so that a deterministic function of a completely deterministic dynamical system provides a model of IID randomness.*

Here are some examples where the parameter is sample size.

**Example 41 (IID Samples)** *Let $X_i$, $i \in \mathbb{N}$ be samples from an IID distribution, and $Z_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ be the sample mean. Then $Z_n$ is a one-parameter stochastic process. The point of the ordinary law of large numbers is to reassure us that $Z_n \to \mathbf{E}[X_n]$ a.s. The point of the central limit theorem is to reassure us that $\sqrt{n}(Z_n - \mathbf{E}[X])$ has constant average size, so that the sampling fluctuation $Z_n - \mathbf{E}[X]$ must be shrinking as $\sqrt{n}$ grows.*

*If $X_i$ is the indicator of a set, this convergence means that the relative frequency with which the set is occupied will converge on its true probability.*

**Example 42 (Non-IID Samples)** *Let $X_i$ be a non-IID one-sided discrete-parameter process, say a Markov chain, and again let $Z_n$ be its sample mean, now often called its "time average". The usual machinery of the law of large numbers and the central limit theorem are now inapplicable, and someone who has just taken 36-752 has, strictly speaking, no idea as to whether or not their time averages will converge on expectations. Under the heading of ergodic theory, we will see when this will happen. Since this is the situation with all interesting time series, the application of statistical methods to situations where we cannot contrive to randomize depends crucially on ergodic considerations.*

**Example 43 (Estimating Distributions)** *Recall Example 10, where we looked at the sequence of empirical distributions $\hat{P}_n$ for samples from an IID datasource. We would like to be able to say that $\hat{P}_n$ converges on $P$. The usual way to do this, if our samples are of a real-valued random variable, is to consider the empirical cumulative distribution function, $F_n$. For each n, this may be regarded as a one-parameter random process ($T = \mathbb{R}$, $\Xi = [0, 1]$), and the difficulty is to show that this sequence of random processes converges to $F$. The usual way is to show that $E_n \equiv \sqrt{n}(F_n - F)$, the* empirical process, *converges to a fixed process, related to the Wiener process, depending only on the true distribution $F$. (See Figure 4.1.) Since the magnitude of the limiting process is constant, and $\sqrt{n}$ grows, it follows that $F_n - F$ must shrink. So theorizing even this elementary bit of statistical inference really requires* two *doses of stochastic process theory, one to get a grip on $F_n$ at each n, and the other to get a grip on what happens to $F_n$ as n grows.*

**Example 44 (Doob's Martingale)** *Let $X$ be a random variable, and $\mathcal{F}_i$, $i \in \mathbb{N}$, a sequence of increasing $\sigma$-algebras (i.e. a filtration). Then $Y_i = \mathbf{E}[X|\mathcal{F}_i]$ is a one-sided discrete-parameter stochastic process, and in fact a martingale. Martingales in general are one-parameter stochastic processes. Note that posterior mean parameter estimates, in Bayesian inference, are instances of Doob's martingale.*

Here are some examples where the one-dimensional parameter is neither time nor sample size.

**Example 45 (The One-Dimensional Ising Model)** *This system serves as a toy model of magnetism in theoretical physics. Atoms sit evenly spaced on the points of a regular, infinite, one-dimensional crystalline lattice. Each atom has a magnetic moment, which is either pointing north ($+1$) or south ($-1$). Atoms are more likely to point north if their neighbors point north, and vice-versa. The natural index here is $\mathbb{Z}$, so the parameter is discrete and two-sided.*

**Example 46 (Text)** *Text (at least in most writing systems!) can be represented by a sequence of discrete values at discrete, ordered locations. Since texts can be arbitrarily long, but they all start somewhere, they are discrete-parameter, one-sided processes. Or, more exactly, once we specify a distribution over sequences from the appropriate alphabet, we will have such a process.*

**Example 47 (Polymer Sequences)** *Similarly, DNA, RNA and proteins are all heteropolymers — compounds in which distinct constituent chemicals (the monomers) are joined in a sequence. Position along the sequence (chromosome, protein) provides the index, and the nature of the monomer at that position the value.*

Linguists believe that no Markovian model (with finitely many states) can capture human language (Chomsky, 1957; Pullum, 1991). Whether this is true of DNA sequences is not known. In both cases, hidden Markov models are used extensively Baldi and Brunak (2001); Charniak (1993), even if they can only be approximately true of language.

## 4.2 Operator Representations of One-Parameter Processes

Consider our favorite discrete-parameter process, say $X_t$. If we try to relate $X_t$ to its *history*, i.e., to the preceding values from the process, we will often get a horribly complicated probabilistic expression. There is however an alternative, which represents the dynamical part of any process as a remarkably simple semi-group of operators.

**Definition 48 (Shift Operators)** *Consider $\Xi^T$, $T = \mathbb{N}, = \mathbb{Z}, = \mathbb{R}^+$ or $= \mathbb{R}$. The* shift-by-$\tau$ *operator $\Sigma_\tau$, $\tau \geq 0$, maps $\Xi^T$ into itself by shifting forward in time: $(\Sigma_\tau)x(t) = x(t + \tau)$. The collection of all shift operators is the* shift semi-group *or* time-evolution semi-group.

(A semi-group does not need to have an identity element, and one which does is technically called a "monoid". No one talks about the shift monoid or time-evolution monoid, however.)

Before we had a $\Xi$-valued stochastic process $X$ on $T$, i.e., our process was a random function from $T$ to $\Xi$. To extract individual random variables, we used the projection operators $\pi_t$, which took $X$ to $X_t$. With the shift operators, we simply have $\pi_t = \pi_0 \circ \Sigma_t$. To represent the passage of time, then, we just apply elements of this semi-group to the function space. Rather than having complicated dynamics which gets us from one value to the next, by working with shifts on function space, all of the complexity is shifted to the initial distribution. This will prove to be extremely useful when we consider stationary processes in the next lecture, and even more useful when, later on, we want to extend the limit theorems from IID sequences to dependent processes.

## 4.3 Exercises

**Exercise 4.1 (Existence of proto-Wiener processes)** *Use Theorem 29 and the properties of Gaussian distributions to show that processes exist which satisfy*

*points (1)–(3) of Example 38 (but not necessarily continuity). You will want to begin by finding a way to write down the FDDs recursively.*

**Exercise 4.2 (Time-Evolution Semi-Group)** *These are all very easy, but worth the practice.*

1. *Verify that the time-evolution semi-group, as described, is a monoid, i.e., that it is closed under composition, that composition is associative, and that there is an identity element. What, in fact, is the identity?*

2. *Can a one-sided process have a shift group, rather than just a semi-group?*

3. *Verify that $\pi_\tau = \pi_0 \circ \Sigma_\tau$.*

4. *Verify that, for a discrete-parameter process, $\Sigma_t = (\Sigma_1)^t$, and so $\Sigma_1$ generates the semi-group. (For this reason it is often abbreviated to $\Sigma$.)*
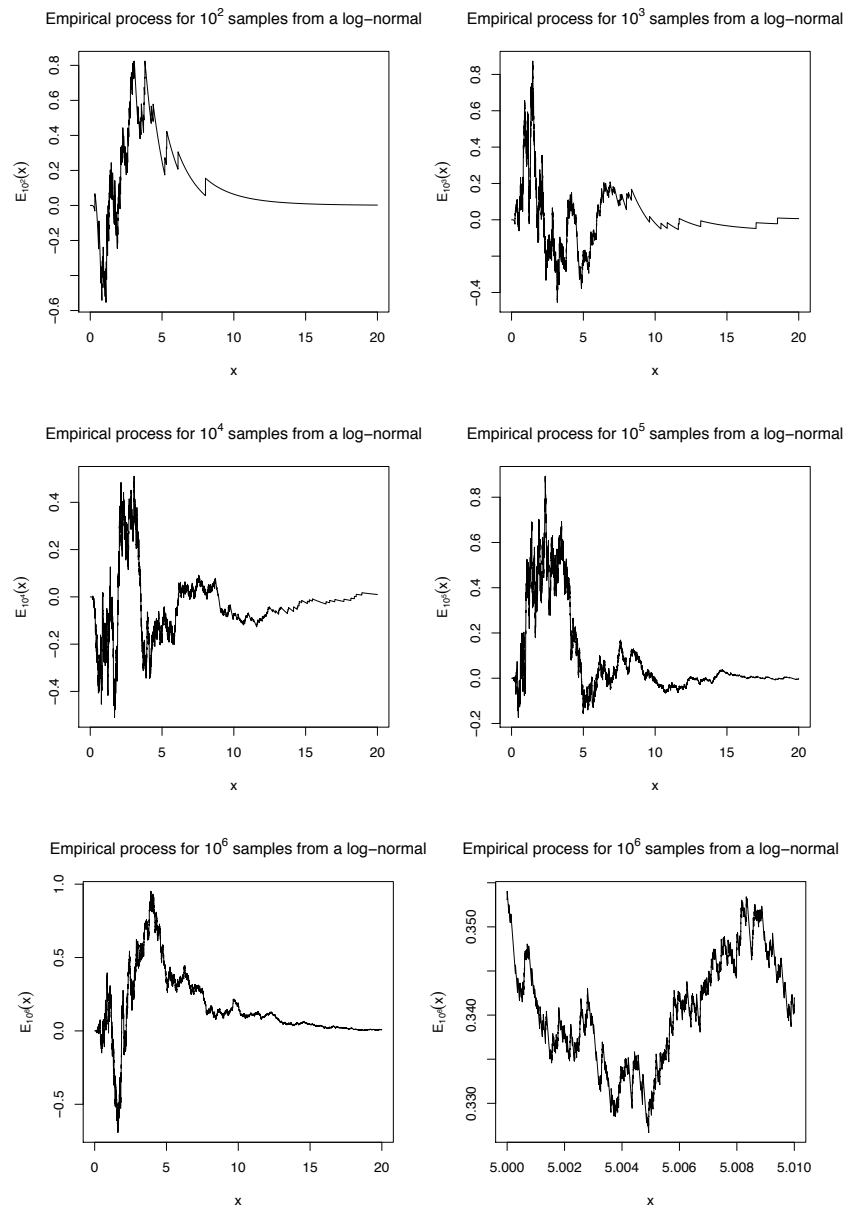
Figure 4.1: Empirical process for successively larger samples from a log-normal distribution, mean of the log = 0.66, standard deviation of the log = 0.66. The samples were formed by taking the first $n$ variates from an over-all sample of 1,000,000 variates. The bottom right figure magnifies a small part of $E_{10^6}$ to show how spiky the paths are.