is exactly given everywhere by a non-linear function of Y_s, $\dot{Y}_s$ computed at a single point s.)

Modeling and Ergodicity

A common physical situation is one for which there is only a single time record of some physical phenomenon. For instance, $f(t)$ might be a record of tidal heights as a function of time over a particular 20 years, or of seismic activity during some chosen period in White Plains, New York, or perhaps of atmospheric turbulence over a certain week at a given location. Of course, there is no possibility of repeating these experiments under the same conditions. How does one restart the earth's tides or geological activity? Yet, even with only a single time track $f(t)$ of the phenomenon available, it is often useful and illuminating to model the phenomenon as a stationary process. How can this be done? What is the conceptual justification or rationale behind this kind of modeling?

Here is the reasoning: We have a function $f(t)$, recorded over some time period that is long compared to the "scale" of the phenomenon. Look at it. Is it "homogeneous" in time? What this means is, can you detect any systematic changes in $f(t)$ over the time period? For instance, does $f(t)$ tend to be larger on the average as time increases, or does it tend to oscillate more rapidly at the beginning of the period? If the *average* properties of $f(t)$ remain roughly the same during the recorded interval, then call it a *time-homogeneous record.*

Thus, by definition, a time-homogeneous record has statistical properties that are the same for all parts of the recorded interval. By statistical properties, I mean the various long-run time averages associated with the record, such as

$$\bar{f} = \frac{1}{T}\int_0^T f(t)\,dt$$

$$\bar{f}^2 = \frac{1}{T}\int_0^T [f(t)]^2\,dt$$

$$\overline{f(t)f(t+\tau)} = \frac{1}{T}\int_0^T f(t)f(t+\tau)\,d\tau.$$

Obviously, there is an infinite number of similar time averages that we could write out. The totality of these averages we refer to as the *statistical properties* of the record $f(t)$.

The transition to a probabilistic model occurs when we decide to *consider all time records having the same statistical properties as the given*

record $f(t)$. Thus, we construct an artificial infinity of different time records, all of which have the same statistical properties. This new universe is sometimes called *the ensemble.* The reasons for passing to the ensemble rather than looking at $f(t)$ alone are simple:

a It is generally impossible to predict future values of a phenomenon exactly over any future time interval. But for many phenomena one can assume with some confidence that a future time record will have the same statistical properties as the observed time record. Hence it will be a member of the ensemble.

b If we are interested only in the statistical properties of the record, then we may as well work with the ensemble. In fact, working with the ensemble allows the theory to be more clearly formulated.

Take all the time records in the ensemble as being defined for $-\infty < t < \infty$. Then this *ensemble forms the outcome space for a stationary process* X_t, $-\infty < t < \infty$. In other words, each outcome of the process is a function in the ensemble. Or you can think of the process as a device or system whose output will be some record in the ensemble. Then we can shake the device, begin it again, and get another member of the ensemble as output.

Now that we know what our outcome space is for the model, there is still the problem of how to assign probabilities. We use again the basic heuristic that stands behind virtually every model in this book:

The probabilities assigned are a summary of the statistical properties of the record.

In coin-tossing, we assigned probability $\frac{1}{2}$ to heads on the grounds that, in the long run, heads turned up half the time. This same reasoning is what we use in construction of stationary models—assign probabilities and expectations by means of long-run time averages. Thus, suppose that $\frac{3}{4}$ of the time $f(t)$ is between 1 and 2. Translate this as the assignment

$$P(1 \leq X_t \leq 2) = \tfrac{3}{4}.$$

If

$$\frac{1}{T}\int_0^T f(t)\,dt = 6$$

translate this into the model as

$$EX_t = 6.$$

We get joint probabilities the same way: Suppose that $\frac{1}{5}$ of the total time the record $f(t)$ satisfies the inequalities

$$0 \leq f(t) \leq 3,\ 1 \leq f(t+1) \leq 2.$$

Then put

$$P(0 \leq X_t \leq 3,\ 1 \leq X_{t+1} \leq 2) = \tfrac{1}{5}.$$

The most important single translation bridge is to compute

$$\varphi(\tau) = \frac{1}{T}\int_0^T (f(t) - \bar{f})(f(t+\tau) - \bar{f})\,dt$$

and consider X_t to be a second-order stationary process with autocorrelation $\Gamma(\tau) = \varphi(\tau)$.

Once we have constructed our model, we want to know that the structure of the model is internally consistent with the frequency heuristic we have used to construct the model. For instance, within a model for independent tosses of a fair coin, we can compute the probabilistic distribution of the proportion of heads in n tosses. What internal consistency means in this example is that since probabilities were assigned in an attempt to construct a model for a phenomenon that produced an experimentally observed proportion of one-half heads, it would be shocking if the distribution of this proportion, computed within the model, was not highly concentrated around $\frac{1}{2}$. Of course, the result that provides internal consistency with the translation heuristic in coin-tossing is the law of large numbers.

The analogous validation for stationary processes is provided by the famous *ergodic theorem*. This theorem states (under a restriction we discuss later), that long-term time averages along any single output function of a stationary process converge to the expectations of the corresponding variables. A particular case of this theorem is the result that the time-average of X_t equals μ, the common mean of the process, or

$$\mu \simeq \frac{1}{T}\int_0^T X_t\,dt.$$

Similarly, if $\mu = 0$,

$$\sigma^2 \simeq \frac{1}{T}\int_0^T X_t^2\,dt,$$

and

$$\Gamma(\tau) \simeq \frac{1}{T}\int_0^T X_{t+\tau}X_t\,dt.$$

Another way of looking at the ergodic theorem is to say that it guarantees that all possible outcome records of the process have the same

statistical properties. Furthermore, it guarantees that these various common values of long-run time averages are equal to the expected values of the corresponding variables. In other words, just as in the case of independence, we have the fundamental result:

$$\textit{long-run time average} = \textit{probabilistic average.}$$

The stationary processes for which long-run time averages equal probabilistic averages are called *ergodic*. Not all stationary processes are ergodic. To see what non-ergodic stationary processes look like, here is a simple, but typical, discrete time example: Consider two coins. Coin I is fair, coin II has probability $\frac{1}{3}$ of heads. The process is gotten by picking either one of the coins with probability $\frac{1}{2}$ and tossing the chosen coin repeatedly. Let $X_1, X_2, \ldots$ be zero or one as the tosses of the chosen coin result in tails or heads. There is no difficulty in checking that this process is stationary. Think of it: Why should probabilities change as you shift the variables? Now notice how the time averages behave: The expression

$$\frac{X_1 + \cdots + X_n}{n}$$

converges to $\frac{1}{2}$ if coin I was chosen, but converges to $\frac{1}{3}$ if coin II was chosen. Each of these possibilities has probability one-half. Thus, there is no single limiting time average. What is happening is that half of the outcomes have the statistical properties of fair coin-tossing, the other half have the statistical properties of biased coin-tossing with $P(H) = \frac{1}{3}$. The probabilistic average is over both of these chunks, i.e.

$$\begin{aligned} EX_1 &= P(H \text{ on 1st toss}) \\ &= P(H \text{ on 1st toss} \mid \text{coin I chosen})P(\text{coin I chosen}) \\ &\quad + P(H \text{ on 1st toss} \mid \text{coin II chosen})P(\text{coin II chosen}) \\ &= \tfrac{1}{2} \cdot \tfrac{1}{2} + \tfrac{1}{3} \cdot \tfrac{1}{2} = \tfrac{5}{14}. \end{aligned}$$

Each of these two chunks, considered by itself, is an ergodic process. You can look at this process as gotten by putting all the time records generated by process I (coin I) in one urn, all the time records generated by process II (coin II) in another urn, and deciding which urn to select from by tossing a fair coin.

This example is typical of the way that every stationary process, if it is not ergodic, is manufactured out of ergodic processes. Conceive of any number of distinct systems which are ergodic in the sense that all outputs of any one system have common statistical properties. Decide on the basis of some random experiment which system's start button to push. Then this defines a stationary process. Obviously the time averages depend on which system you started, but the probabilistic average is

over all systems. Hence the process is not ergodic. This is a pretty general description of the structure of non-ergodic processes. What it shows basically is that every non-ergodic stationary process is really a sort of lumping together of non-interacting ergodic processes. If you come up in your modeling with a non-ergodic process you are usually doing something wrong—you have not completely decomposed the system down into its separate component parts.

How can we look at a stationary process and check that it is ergodic? If we have one long time record, think of cutting it up into N pieces. If the process is such that variables in it which are separated by a long-time stretch are nearly independent, then these N observation intervals can be considered as N nearly independent records of the process. In such a situation, a time average over the entire record period is almost like taking the average of N independent variables. The law of large numbers gives the result that this average is about equal to the expectation. This gives a rudimentary proof of the fact: *If for a stationary process the variables in two widely separated time intervals are nearly independent, then probabilistic averages (expectations) are equal to the limiting time averages.* In other words, *independence at widely separated time points implies that the process is ergodic.*

What processes have this property? Certainly a periodic process

$$X_t = V \sin \lambda t + U \cos \lambda t$$

does not! If you know X_t over any small time interval, then you know exactly its values over any other time interval no matter how far away. And, in fact, this process is extremely non-ergodic. It is known for Gaussian processes that they are ergodic if and only if they have no periodic components, that is, if $F(d\lambda)$ does not assign positive mass to any single points. In particular, this holds if

$$\Gamma(\tau) \to 0 \quad \text{as} \quad |\tau| \to \infty.$$

You can see by looking at the definition of stability for Markov processes, that a stationary Markov process with stable transition probabilities ought to have this sort of asymptotic independence. Indeed it does, and these processes are also ergodic.

Problem 53 Show directly that a periodic process

$$X_t = V \sin \lambda t + U \cos \lambda t$$

is not ergodic by computing the time averages of X_t^2 and $X_t X_{t+\tau}$ over $[0, T]$ and showing that these do not converge as $T \to \infty$ to EX_t^2, $\Gamma(\tau)$.