Chapter 1

Introduction

blah, blah, blah

Chapter 2

Modelling Sequences

2.1 Random Processes

Consider, in general terms, a random process R that generates a sequence of variables, $R_1, R_2, R_3, \ldots$, where the index i gives the discrete *stage* in the sequence, and each variable R_i randomly takes a value $r_i \in \mathcal{R}$. Then, for some arbitrary sequence length n, we define

$$\vec{\mathbf{R}}_n \quad = \quad (R_1, R_2, \ldots, R_n) \tag{2.1}$$

to be a length-n sequence of random variables, i.e. $|\vec{\mathbf{R}}_n| = n$, and further define

$$\vec{\mathbf{r}}_n \quad = \quad (r_1, r_2, \ldots, r_n) \in \mathcal{R}^n \tag{2.2}$$

to be a corresponding length-n sequence of values. The probability (for a discrete-value process) or probability density (for a continuous-value process) of a given sequence $\vec{\mathbf{r}}_n$ is then defined as

$$P(\vec{\mathbf{R}}_n = \vec{\mathbf{r}}_n) \quad = \quad P(R_1 = r_1, \ldots, R_n = r_n) \; = \; p(r_1, \ldots, r_n)\,. \tag{2.3}$$

Hence, note that if R is a discrete-value process then we must have

$$\sum_{\vec{\mathbf{r}}_n \in \mathcal{R}^n} P(\vec{\mathbf{R}}_n = \vec{\mathbf{r}}_n) \quad = \quad \sum_{r_1 \in \mathcal{R}} \cdots \sum_{r_n \in \mathcal{R}} p(r_1, \ldots, r_n) \quad = \; 1\,. \tag{2.4}$$

Alternatively, if R is a continuous-value process, then we must similarly have

$$\int_{\mathcal{R}^n} P(\vec{\mathbf{R}}_n = \vec{\mathbf{r}}_n)\, d\vec{\mathbf{r}}_n \quad = \quad \int_{\mathcal{R}} \cdots \int_{\mathcal{R}} p(r_1, \ldots, r_n)\, dr_1 \cdots dr_n \; = \; 1\,. \tag{2.5}$$

In other words, given the sequence length n, the set $\mathcal{R}^n$ of all possible sequences $\vec{\mathbf{r}}_n$ covers the entire probability space.

This latter property causes modelling problems if we do not know in advance the exact length of a sequence; for example, the set of all length-1 and length-2 sequences already covers twice the entire probability space. In practice, suppose we have observed a given sequence $\vec{\mathbf{r}}_n$. How do we know if the underlying process R has terminated, or will instead continue to generate another observed value r_{n+1}, leading to the extended sequence $\vec{\mathbf{r}}_{n+1}$? Similarly, how do we know that the first observed value r_1 was not in fact part of a longer, unobserved sequence of values $\ldots, r_{-2}, r_{-1}, r_0$?

In order to handle such difficulties, we distinguish between a so-called *incomplete* sequence $\vec{\mathbf{r}}_n$, and a *complete* sequence $<\vec{\mathbf{r}}_n>$ that has definite stages of initiation and termination. Thus, we see that each length-2 incomplete sequence starts with a length-1 incomplete sequence, so that measuring the set of all length-1 and length-2 incomplete sequences amounts to double counting.

A complete sequence may be specified by introducing indicator variables that define the start and end of the sequence. Thus, indicator I_0 that takes a value of 1 if the sequence starts at stage 0 (i.e. just prior to value r_1 in the complete sequence), or a value of 0 if it does not. Similarly, indicator T_{n+1} that takes a value of 1 if the sequence terminates at stage $n+1$ (i.e. just after value r_n in the complete sequence), or a value of 0 if it does not. As one consequence, we may now also consider the two *partially complete* sequences, namely $<\vec{\mathbf{r}}_n$, which was initiated at stage 0 but not yet terminated (i.e. $I_0 = 1$ but $T_{n+1} = 0$), and $\vec{\mathbf{r}}_n>$, which was terminated at stage $n+1$ but not initiated at stage 0 (i.e. $T_{n+1} = 1$ but $I_0 = 0$). As another consequence, these definitions of the indicator variables permit empty, or length-0, sequences. If this is not desirable, then we could replace indicators I_0 and T_{n+1} by new indicators I'_1 and T'_n, respectively. Then I'_1 takes the value 1 (or 0) if r_1 is (or is not) the first observation in a complete sequence, and T'_n takes the value 1 (or 0) if r_n is (or is not) the final observation in the complete sequence.

The probability or probability density of a given complete sequence $<\vec{\mathbf{r}}_n>$ is now defined as

$$P(<\vec{\mathbf{r}}_n>) \quad = \quad P(I_0 = 1, R_1 = r_1 \ldots, R_n = r_n, T_{n+1} = 1)\,. \tag{2.6}$$

Hence, for a discrete-value process we now obtain

$$\sum_{\vec{\mathbf{r}}_n \in \mathcal{R}^n} P(<\vec{\mathbf{r}}_n>) \quad = \quad P(I_0 = 1, T_{n+1} = 1) \;\doteq\; P(N = n)\,, \tag{2.7}$$

where we have introduced the random variable N to denote the length of an arbitrary complete sequence. For the corresponding continuous-value process, we likewise deduce that

$$\int_{\mathcal{R}^n} P(<\vec{\mathbf{r}}_n>)\, d\,\vec{\mathbf{r}}_n \quad = \quad P(I_0 = 1, T_{n+1} = 1) \;\doteq\; P(N = n)\,. \tag{2.8}$$

We therefore deduce that the set $\mathcal{R}^* = \bigcup_{n=0}^{\infty} \mathcal{R}^n$ of all complete sequences of arbitrary length covers the entire probability space exactly once, since for the discrete-value case we have

$$\sum_{<\vec{\mathbf{r}}_*>\in\mathcal{R}^*} P(<\vec{\mathbf{r}}_*>) \;=\; \sum_{n=0}^{\infty}\sum_{\vec{\mathbf{r}}_n\in\mathcal{R}^n} P(<\vec{\mathbf{r}}_n>)$$
$$= \sum_{n=0}^{\infty} P(N=n) \;=\; 1\,, \tag{2.9}$$

and for the continuous-value case we have

$$\int_{\mathcal{R}^*} P(<\vec{\mathbf{r}}_*>)\,d\vec{\mathbf{r}}_* \;=\; \sum_{n=0}^{\infty}\int_{\mathcal{R}^n} P(<\vec{\mathbf{r}}_n>)\,d\vec{\mathbf{r}}_n$$
$$= \sum_{n=0}^{\infty} P(N=n) \;=\; 1\,. \tag{2.10}$$

2.2 Markov Processes

In the previous section we defined random processes and the sequences they generate. It is generally assumed that a random process R obeys strict causality, such that the values generated by the process form a *temporal* sequence, where the distribution of values for variable R_i, at stage i, depends only upon the values generated previously in the sequence at stages $i-1, i-2, \ldots, 1$. In addition, for a complete sequence the distribution of the variable R_1 at the initial stage depends strongly upon being first in the sequence, and likewise the distribution of the variable R_n, for a length-n sequence, depends strongly upon both the past values in the sequence and on the fact that it is the final stage.

Causality therefore indicates that a probability model of the sequences generated by a random process R can be factored in terms of conditioning, at each stage i, the random variable R_i on the values of the previous variables $R_{i-1}, R_{i-2}, \ldots, R_1$. Thus, for any given complete sequence $<\vec{\mathbf{r}}_n>$, we obtain the model

$$\begin{aligned} P(<\vec{\mathbf{r}}_n>) \;&=\; P(I_0=1)\,P(R_1=r_1 \mid I_0=1) \\ &\quad P(R_2=r_2 \mid R_1=r_1, I_0=1)\cdots \\ &\quad P(R_n=r_n \mid R_{n-1}=r_{n-1},\ldots,R_1=r_1,I_0=1) \\ &\quad P(T_{n+1}=1 \mid R_n=r_n,\ldots,R_1=r_1,I_0=1) &(2.11)\\ &=\; P(I_0=1)\prod_{i=1}^{n} P(R_i=r_i \mid \vec{\mathbf{R}}_{i-1}=\vec{\mathbf{r}}_{i-1}, I_0=1) \\ &\quad P(T_{n+1}=1 \mid \vec{\mathbf{R}}_n=\vec{\mathbf{r}}_n, I_0=1)\,. &(2.12)\end{aligned}$$

Observe that we may adjust this model to allow for partially complete or incomplete sequences by modifying the corresponding boundary conditions I_0 and T_{n+1}.

In practice, this factored model is usually simplified further by limiting the conditional dependency on past values to a small number. In particular, one might assume the Markov property, such that the value at any stage i depends only upon the immediate past value at stage $i-1$, namely

$$\begin{aligned} P(<\vec{\mathbf{r}}_n>) &= P(I_0=1)\,P(R_1=r_1 \mid I_0=1) \\ & \quad \prod_{i=2}^{n} P(R_i=r_i \mid R_{i-1}=r_{i-1}) \\ & \quad P(T_{n+1}=1 \mid R_n=r_n)\,. \end{aligned} \tag{2.13}$$

Alternatively, this 1st order Markov model may be generalised to an m-th order model by limiting the dependency of R_i to only the previous m values, namely

$$\begin{aligned} & P(R_i=r_i \mid \vec{\mathbf{R}}_{i-1}=\vec{\mathbf{r}}_{i-1}, I_0=1) \\ & \quad \doteq \begin{cases} P(R_i=r_i \mid \vec{\mathbf{R}}_{i-m,i-1}=\vec{\mathbf{r}}_{i-m,i-1}) & \text{if } i \geq m+1\,, \\ P(R_i=r_i \mid \vec{\mathbf{R}}_{1,i-1}=\vec{\mathbf{r}}_{1,i-1}, I_0=1) & \text{if } i \leq m\,. \end{cases} \end{aligned} \tag{2.14}$$

where

$$\vec{\mathbf{R}}_{i,j} = (R_i, R_{i+1}, \ldots, R_j)\,. \tag{2.15}$$