3. Recurrence Relations With Noise: Markov Chains.

There are many situations where we lack the complete information needed to predict the future. One might model such things by

$$x_n = f(x_{n-1}, \xi_n)$$

Here the function f is known. We know the current state x_{n-1}, but in order to say what x_n is, we need to know the value of ξ_n. If we lack that knowledge we cannot say what x_n will be. If we iterate as before we get

$$x_n = f_n(x_0 : \xi_1, \ldots, \xi_n)$$

where x_n depends on the initial state and the unknown variables $\xi_1, \ldots, \xi_n$. The variables ξ are often termed "Noise".

Examples:

1. The price of a stock from one period to the next is related by

$$x_n = x_{n-1}(1 + R_n)$$

where R_n is the return for the n-th period.

2. A gambler's cash on hand after n games of poker is given by

$$x_n = x_{n-1} + \xi_n$$

where ξ_n are his winnings (or loss) in the n-th game.

While it may not be possible to predict the value of the noise term ahead of time it may often be possible to say some thing of a statistical nature about it.

3. In Example 2, suppose that the gambling game is much simpler than poker and consists of tossing a coin with the gambler winning a dollar on heads and losing a dollar on tails. Then each ξ_n is ± 1. Assuming the coin is fair and the successive tosses are "independent" each ξ takes the value ± 1 with probability $\frac{1}{2}$. While it is not possible to make deterministic statements about the winnings x_n it is possible to say that

$$P[x_n = x_{n-1} + 1] = P[x_n = x_{n-1} - 1] = \frac{1}{2}$$

or that the winnings go up ot down with probability $\frac{1}{2}$. Independence here is in the statistical sense, i.e.

$$P[\xi_j = \epsilon_j : 1 \le j \le n] = \frac{1}{2^n}$$

for every choice of $\epsilon_j = \pm 1$. In such models it is possible to make staistical or probabilistic statements and we can calculate

$$P[x_n \le y | x_0 = x] = F_n(x, y) = P[\xi_1 + \ldots + \xi_n \le y - x]$$

4*. Obtain an explicit formula (in terms of binomial coefficients) for the probability in Example 3. Can you compute $E[x_n|x_0]$ and $Var[x_n|x_0]$ as functions of x_0 and n?

The statistical independence of the successive noise terms $\xi_1, \ldots, \xi_n$ results in the Markov property for the observed sequence $x_1, \ldots, x_n$. There are different ways of describing the Markov property. If we have a finite sequence of quantities $x_0, \ldots, x_n$ that are possibly random, then such a sequence is said to have the Markov Property if

$$P[x_{k+1} \leq y|x_0, \ldots, x_k] = P[x_{k+1} \leq y|x_k]$$

for all y and $1 \leq k \leq n-1$. In words to make a probabilistic statement about x_{k+1} from a knowledge of $x_1, \ldots, x_k$ we need to know only x_k. The conditional probabilities depend only on x_k the value at the last time k. There is a lack of memory. The probabilities of future evolution given the history so far, depend only on the current state. To see why the independence of the noise leads to the Markov property let us make a calculation. Assume that all values are discrete so that the conditional probabilities can be calculated as ratios

$$(1) \qquad P[x_{k+1} = a_{k+1}|x_0 = a_0, \ldots, x_k = a_k] = \frac{P[x_0 = a_0, \ldots, x_{k+1} = a_{k+1}]}{P[x_0 = a_0, \ldots, x_k = a_k]}$$

We use the realtion $x_{k+1} = f(x_k, \xi_{k+1})$ to define the set

$$E_{a_k, a_{k+1}} = [\xi : f(a_k, \xi) = a_{k+1}]$$

Because ξ_{k+1} is independent of $\xi_1 \ldots, \xi_k$, and $x_0, \ldots, x_k$ depend only on $\xi_1, \ldots, \xi_k$,

$$\begin{aligned} P[x_0 = a_0, \ldots, x_{k+1} = a_{k+1}] &= P[\{x_0 = a_0, \ldots, x_k = a_k\} \cap E_{a_k, a_{k+1}}] \\ &= P[x_0 = a_0, \ldots, x_k = a_k] P[E_{a_k, a_{k+1}}] \end{aligned}$$

and the conditional probability (1) reduces to

$$P[\xi_{k+1} \in E_{a_k, a_{k+1}}] = P[x_{k+1} = a_{k+1}|x_k = a_k]$$

which is the Markov property.

The statistical independence of the successive noise terms is crucial for the Markov property. For instance if have a model like $x_{k+1} = x_k + \xi_{k+1}$ with $\xi_k \equiv \xi$ so that the noise is chosen once and fixed, then if we know x_0 and x_i, then $x_2 = x_1 + (x_1 - x_0) = 2x_1 - x_0$. This means, given x_i a knowledge of x_0 is still helpful in predicting x_2. The Markov property is not valid in this case.

Examples:

5. Consider an urn containing 10 white and 10 red balls. A ball is drwan at random and is returned to the urn after a second ball is drawn, which is returned after drawing the third ball and so on. Each $x_k = W$ or R depending on the color of the ball. Cleraly in this case

$$\begin{aligned} P[x_{k+1} = W|x_0, \ldots, x_k] &= \frac{9}{19} \quad \text{if} \quad x_k = W \\ P[x_{k+1} = W|x_0, \ldots, x_k] &= \frac{10}{19} \quad \text{if} \quad x_k = B \end{aligned}$$

and the Markov property holds. We can easily construct independent random variables $\{\xi_j\}$ such that $x_{k+1} = f(x_k, \xi_{k+1})$.

Let us define $\{\xi_j\}$ to be random variables that take the values $1, 2. \ldots, 19$ with equal probabilities of $\frac{1}{19}$. We can define

$$\begin{aligned} f(W, r) &= W \quad \text{if} \quad r = 1, \ldots, 9 \\ f(W, r) &= B \quad \text{if} \quad r = 10, \ldots, 19 \\ f(B, r) &= B \quad \text{if} \quad r = 1, \ldots, 9 \\ f(B, r) &= W \quad \text{if} \quad r = 10, \ldots, 19 \end{aligned}$$

This clearly works as will many other represntations. The model cannot be verified by observing $\{x_j\}$.

Example:

6.* If the joint probability density of $x_1, x_2 \ldots, x_n$ given x_0 is given by

$$\Pi_{i=0}^{n-1} f(x_i, x_{i+1})$$

where

$$f(x, y) = \frac{1}{\sqrt{2\pi a(x)}} \exp[-\frac{(y - b(x))^2}{2a(x)}]$$

show that the sequence has the Markov property and find a representation in terms of independent random variables $\{\xi_j\}$. Here $a(x)$ and $b(x)$ are nice functions of x with $a(x) > 0$.

7. A special case is when $x_{k+1} = x_k e^{R_j}$, used in financial modeling. Here $\{R_j\}$ are single period log-returns that are usually assumed to be mutually independent and Gaussian with some common mean μ and variance σ^2. This leads to a joint density involving the log-normal densities.

$$f(x_1, \ldots, x_n | x_0) = \Pi_{i=0}^{n-1} f(x_i, x_{i+1})$$

with

$$f(x, y) = \frac{1}{y} \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{(\log \frac{y}{x} - \mu)^2}{2\sigma^2}]$$

While a representation of the form $x_1 = f(x_0, \xi_1)$ does not allow us to say anything definite about x_1 from a knowledge of the model and the starting position x_0, as we saw before we can make probabilistic statements concerning x_1 if we know the probability distribution of ξ_1. We can compute expecations $E[g(x_1)|x_0]$ by averaging with respect to the distribution of ξ_1. This will turn out to be important.

Examples:

8. Suppose an asset has a random income stream as well as a random growth rate. If we denote by ξ_n the income during the n-th period and r_n the random rate of return for the

n-th period then the asset A_{n+1} at time $n+1$ is given by $A_{n+1} = A_n\, r_n + \xi_n$. We need to know both ξ_n and r_n to compute A_{n+1} from A_n. So the "noise" here has two components. In a general model there is of course no restriction on the number different sources of the noise that determine the change in the sysytem.

9. In an intersting question is to go from a discrete time model (recurrence relation) with noise to a an ODE with noise. While it is possible to define a continuous analog of the random return model by

$$dA(t) = A(t)r(t)dt$$

where $r(t)$ is the instantaneous random rate of return, it is not so easy to determine what the distribution of $r(t)$ should be. In the discrete world it is natural to assume that $r = e^R$ with R having a Gaussian distribution with mean μ and variance σ^2. If the duration of time is h units, it is natural to suppose that $R(h) = \int_t^{t+h} r(s)ds$ is Gaussian with mean μh and variance $\sigma^2 h$. If we now wish to guess what the distribution of the instantaneous noise $r(t)$ is, since $r(t) \simeq \frac{R(h)}{h}$, we end up with a Gaussian with mean μ and variance $\frac{\sigma^2}{h}$. In other words $r(t)$ wants to have an infinite variance. This is a problem!

10. One way to solve the problem is not to talk about $r(t)$ but only talk about $z(t) = \int_0^t r(s)ds$. As we saw earlier it is natural to suppose that for $t \geq s \geq 0$, $z(t) - z(s)$ is Gaussian with mean $(t-s)\mu$ and variance $(t-s)\sigma^2$ and that the increments $z(t_j) - z(s_j)$ are independent gaussians if the intervals $[s_j, t_j]$ are disjoint, i.e. do not overlap. Such a family of random variables $z(t)$ can be represented as $\mu t + \sigma\beta(t)$, where $\beta(t)$ is just like the old $z(t)$, but has been normalized to have $\mu = 1$ and $\sigma^2 = 1$. Now, may be our model can be

$$dA(t) = A(t)[\mu + \sigma\beta'(t)]dt = A(t)[\mu dt + \sigma d\beta(t)] \tag{2}$$

This unfortunately is not quite Kosher. The problem is β is not differentiable. After all $\beta'(t) = r(t)$ wants to have an infinite variance. Well we could try to reperesent the ODE in an integrated form. This does not help either, because to make sense out of

$$A(t) = A(0) + \mu \int_0^t A(s)ds + \sigma \int_0^t A(s)d\beta(s)$$

$\beta(\cdot)$ needs to be of bounded variation and it is not. To make progress one has to tackle $d\beta$. This is in fact the Black-Scholes model and we will return to it later.