Lecture 1 - Mathematical Foundations of Stochastic Processes

Charles R. Doering; notes by Tom Beucler & Andre Souza

June 15, 2015

We will talk about stochastic processes in general but focusing on differential equations with white noise coefficients in them. We will try to give an intuitive feeling for this field. There is actually some sense in the whole thing. By the end of the lectures you should have a clue of how to model systems with white noise.

When we are dealing with many-body nonlinear systems, there are too many variables. They are too high dimensional to extract useful information, forcing us to use reduced models. In applied math there is a great tradition of getting reduced models from large or small parameters in equations or a separation of scales in time or length. We also model things with noise. How do we know the difference between random and complicated? Well we don't. This is a difficult concept.

The great victory of this approach is statistical mechanics. Settling for bulk quantities of interest as opposed to 10^{23} degrees of freedom was tractable and has stimulated people in the physics community to try to take this concept to the extreme. There are a lot of specific successes and examples, but no general theory.

Specifically what we be talking about:

1. Markov diffusion processes
2. Brownian motion
3. Gaussian white noise
4. Stochastic differential equations
5. Fokker-Planck equations or Forward Kolmogorov equations
6. Mean first passage times

SDEs are what you write down and the Fokker-Planck equations are what you can actually solve. The latter equations give you the evolution for a probability density function.

A random variable X is characterized by its cumulative distribution function (CDF) which is the probability that a random variable is below some scalar. In symbols ($\mathbb{P}(X \leq x)$). It is a monotonically increasing function with values between 0 and 1. It is "continuous from the right". The probablity distribution function is the derivative of the probability with respect to x.

$$\text{PDF} = \frac{d\text{CDF}}{dx}$$

Another way to say it is PDF $= \mathbb{P}(x \leq X \leq x + dx)/dx$ where dx is an infinitesimal or

$$\mathbb{P}(a < x < b) = \int_a^b \text{PDF}(x)dx$$

Stochastic Processes are a random functions of an "index set" which we will call time.

$$X(t) = \text{Random variable}$$

We can plot $X(t)$ (for a given realization of the random variable), which we will assume to be continuous. We can ask the question "what is the probability that $X(t)$" falls in some window. We will now introduce some notation

$$\text{PDF of } X(t) = \rho(x, t)$$

This is not enough to answer all the statistical questions that we would like to pose. We also need the joint distribution functions ($\rho(x_1, t_1; x_2, t_2)$ 2-time) and $\rho(x_n, t_n; x_{n-1}, t_{n-1}; ...; x_1, t_1)$ n-time. A property of white noise is that $\rho(x_1, t_1)\rho(x_2, t_2)$. This type of process does not have enough structure for us to do modeling. We don't need an uncountable number of joint distributions to have a well-defined probability space. Smooth things tend to have a memory associated with them, thus white noise won't be continuous. The condition

$$1 = \int_{-\infty}^{\infty} \rho(x, t)dx$$

says that "I exist" and the compatibility condition is

$$\rho(x_1, t_1) = \int_{-\infty}^{\infty} \rho(x_1, t_1; x_2, t_2)dx_2.$$

This says that the probability of going through one window is the same as going through the same window as well as an infinitely large window. See Figure 1.

To answer sensible questions about the random process we need all of the joint probability functions. Suppose that we just give the joint probabilities, then we need to check the compatibility conditions.

$$\rho(x_1, t_1) \tag{1}$$
$$\rho(x_1, t_1; x_2, t_2) \tag{2}$$
$$\rho(x_1, t_1; x_2, t_2, ; x_3, t_3) \tag{3}$$
$$\vdots \tag{4}$$

We need to define expectations, averages, and moments of the random variables. The expectation of a random variable is

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x\rho(x)dx = \langle x \rangle = \bar{X}.$$

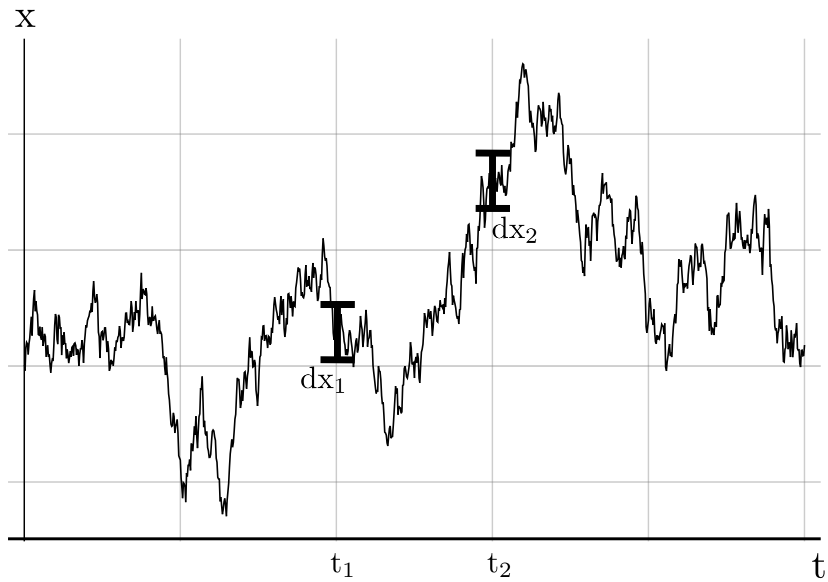

Figure 1: $\rho(x_1, x_2; t_1, t_2)dx_1dx_2$ is the probability that the process $X(t)$ passes through windows of sizes dx_1 and dx_2 at times t_1 and t_2 respectively.

We can also calculate

$$\mathbb{E}(f(X)) = \int_{-\infty}^{\infty} f(x)\rho(x)dx$$

and get moments

$$\mathbb{E}(X^N) = \int_{-\infty}^{\infty} x^N \rho(x)dx.$$

Furthermore we have

$$\mathbb{E}\left(\prod_{j=1}^{N} X(t_j)\right) = \int \cdots \int \prod_{j=1}^{N} x_j \rho(x_1, t_1; ...; x_N, t_N)dx_1 \cdots dx_N$$

which are known as the n-point correlation functions. The two point correlation function is sometimes known as THE correlation function. The moments DO NOT determine the probability distribution (in general). If it is the case that

$$\mathbb{E}(X(t)X(s)) = \mathbb{E}(X(t))\mathbb{E}(X(s))$$

we say the variables are uncorrelated. This does not imply independence, but independence does imply uncorrelated.

One more thing. Let's talk about this idea of independence. Suppose that two events A and B happen. We can look at $\mathbb{P}(A), \mathbb{P}(B)$ and $\mathbb{P}(A \cap B)$ and $\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$. If the variables are independent then $\mathbb{P}(A|B) = \mathbb{P}(A)$. For stochastic processes we may want to know things like "given that my random variable went through window 1, what is the probability that it goes through window 2?" We write this as $\rho(x_2, t_2|x_1, t_1) = \rho(x_2, t_2; x_1, t_1)/\rho(x_1, t_1)$.

We have all these joint probability functions

$$\rho(x_n, t_n|x_{n-1}, t_{n-1}; ...; x_1, t_1) \equiv \frac{\rho(x_n, t_n; x_{n-1}, t_{n-1}; ...; x_1, t_1)}{\rho(x_{n-1}, t_{n-1}; ...; x_1, t_1)}$$

which is the probability of going through my latest window given that I went through all the other windows. We are now in a position to define Markov processes. If

$$\rho(x_n, t_n|x_{n-1}, t_{n-1}; ...; x_1, t_1) = \rho(x_n, t_n|x_{n-1}, t_{n-1}),$$

we can reconstruct the n-point distribution function

$$\rho(x_n, t_n; x_{n-1}, t_{n-1}; ...; x_1, t_1) = \rho(x_1, t_1) \prod_{j=2}^{j=N} \rho(x_j, t_j|x_{j-1}, t_{j-1}).$$

A Markov process is independent of the past, given the present. An example of a Markov process is a first order ODE.

Brownian motion (which is the same as Wiener process) is our next topic. These are random functions of time denoted by

$$W(t) = W_t.$$

The probability density at $t = 0$ is a delta function, $\rho(\omega, 0) = \delta(\omega)$. The transition density is

$$\rho(\omega, t|\omega', t') = \frac{1}{\sqrt{2\pi(t-t')}} e^{-\frac{1}{2}\frac{(w-w')^2}{t-t'}}$$

We have

$$\rho(\omega, t) = \int_{-\infty}^{\infty} \rho(\omega, t|\omega', 0)\rho(\omega', 0)d\omega' = \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2}\frac{\omega'^2}{t}}$$

There are some properties of Brownian motion that are absolutely essential to understanding white noise which we will talk about next time.