Lecture 9 – Markov Processes

In the last lecture, we simplified the joint density $f(\underline{x};\underline{t})$ of a stochastic process using a Gaussian approximation. The other way we can simplify the joint PDF is using time.

A. Conditional Independence

The random vector for a stochastic process is not just any random vector, the components of the vector are ordered in time! So, the dynamics of the process could be used to make a simplification.

One way to think about this is to use conditional probability to try to construct the joint pdf. (Drop t's for shorter formula)

$$\underbrace{f(x_1,\dots,x_n)}_{\text{joint pdf}} = \underbrace{f(x_n \mid x_1,\dots,x_{n-1})}_{\text{conditional pdf}}\,\underbrace{f(x_1,\dots x_{n-1})}_{\text{marginal pdf}}$$

Remember there is a chain rule that lets us build the joint PDF,

$$f(x_1,\dots,x_n) = f(x_1)\, f(x_2 \mid x_1)\, f(x_3 \mid x_1, x_2) \dots f(x_n \mid x_1,\dots,x_{n-1})$$

A markov process has the property of conditional independence. This means that the conditional pdf only depends on the immediate past state.

$$f(x_n \mid x_1, x_2,\dots,x_{n-1}) = f(x_n \mid x_{n-1})$$

$\Pi_{n,n-1}$ special conditional pdf, the transition probability from t_{n-1} to t_n

"memoryless" – future only depends on present, not past

Substituting this property above allows us to simplify the joint pdf,

$$f(x_1, x_2,\dots,x_n) = f(x_1)\, f(x_2 \mid x_1)\, f(x_3 \mid x_2) \dots f(x_n \mid x_{n-1})$$

$$= \underbrace{f(x_1)}_{\text{initial state}} \prod_{i=2}^{n} \underbrace{f(x_i \mid x_{i-1})}_{\pi_{i,i-1}}$$

This can still be complex (every $\pi_{i,i-1}$ can be different) but it is much simpler than before! If we know $\pi_{i,i-1}$, we know everything!

B. Chapman-Kolmogorov Equation

(i) Key concept: propagation in time

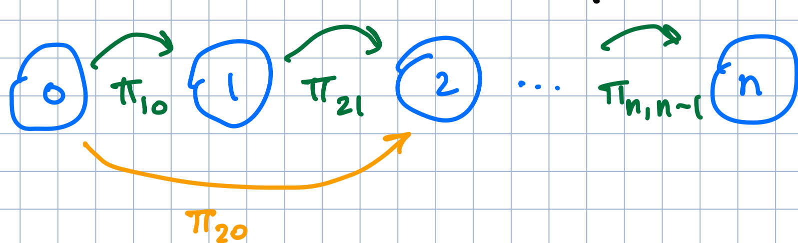
OK, so for a Markov process, if we know $\pi_{i,i-1}$ we can know the entire process statistics. But thinking about just the joint PDF misses something important. By switching to the perspective of using conditional probabilities we adopt a mental model of propagation in time!

- Before: What is the joint distribution in time?
- Now: Given the state at time t, how do I propagate forward using $\pi_{i,i-1}$?

Even if the system is not Markovian, I can still use this perspective (but the system will have a memory).

(ii) Derivation of the Chapman-Kolmogorov Equation

OK, with the concept of propagation in mind, how do we "move" from one state to another? To understand this, consider the state space of a discrete stochastic process



Suppose I know the PDF at the initial state, $f(\underbrace{x_0}_{x_0 = x(t_0)})$. I can get the PDF of the next state using the joint PDF

$$f(x_0, x_1) = f(x_0)\, f(x_1 \mid x_0)$$

$$f(x_1) = \int_{-\infty}^{\infty} f(x_0, x_1)\, dx_0 = \int_{-\infty}^{\infty} \underbrace{f(x_1 \mid x_0)}_{\pi_{10}}\, f(x_0)\, dx_0$$

Law of Total Probability

Of course, I can also do this to step from 1 to 2 and on to step n.

$$f(x_2) = \int_{-\infty}^{\infty} \pi_{21}\, f(x_1)\, dx_1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \pi_{21}\pi_{10}\, f(x_0)\, dx_1\, dx_0 \quad (*)$$

$$\vdots$$

$$f(x_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \pi_{n,n-1} \cdots \pi_{21}\pi_{10}\, f(x_0)\, dx_{n-1} \cdots dx_1\, dx_0$$

What if I want to jump more than one state at a time? I could also build multi-step transition probabilities. Start from $(*)$,

$$f(x_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \pi_{21}\pi_{10}\, f(x_0)\, dx_1\, dx_0 = \int_{-\infty}^{\infty} \underbrace{\left[\int_{-\infty}^{\infty} \pi_{21}\pi_{10}\, dx_1\right]}_{\pi_{20} = f(x_2 \mid x_0)} f(x_0)\, dx_0$$

$$f(x_2) = \int_{-\infty}^{\infty} \pi_{20}\, f(x_0)\, dx_0$$

$$\underbrace{f(x_2 \mid x_0)}_{\pi_{20}} = \int_{-\infty}^{\infty} \underbrace{f(x_2 \mid x_1)}_{\pi_{21}}\, \underbrace{f(x_1 \mid x_0)}_{\pi_{10}}\, dx_1 \qquad (**)$$

Equation $(**)$ is important. It is called the Chapman-Kolmogorov Equation. It tells us how the transition probabilities combine to propagate forward multiple steps. It tells us the dynamics of the transition probabilities.

C. Fokker-Planck Equation

Now we see how we can propagate the marginal densities and transition probabilities. But how do we get the transition probabilities in the first place? We are going to use the above properties of Markov processes to write a differential equation for the transition probabilities, $\pi_{i,i-1}$. Then we can solve this DE to find $\pi_{i,i-1}$! This DE is called a Fokker-Planck equation.

(i) Derivation of F-P

We start w/ Chapman-Kolmogorov (how $\pi_{i,i-1}$ propagates)

$$f(x_2|x_0) = \int_{-\infty}^{\infty} f(x_2|x_1) f(x_1, x_0)\, dx_1$$

$$f(x;t+\Delta t|x_0;t_0) = \int_{-\infty}^{\infty} \underbrace{f(x;t+\Delta t|x-\Delta x;t)}_{\pi_{\Delta x,\Delta t}\text{: short-time kernel}} f(x-\Delta x;t|x_0;t_0)\, \underbrace{d(\Delta x)}_{\text{change of variable}}$$

$$⓪ \xrightarrow{\pi_{10}} ① \xrightarrow{\pi_{21}} ②$$

x_0, t_0 | $x_1 = x-\Delta x$, t | $x_2 = x$, $t+\Delta t$

$dx_1 = d(x-\Delta x) = -d(\Delta x)$

(swap limits, no diff)

Taylor series in time (LHS)

$$f(x;t+\Delta t|x_0) = f(x;t|x_0) + \Delta t \frac{\partial}{\partial t} f(x;t|x_0) + \mathcal{O}(\Delta t^2)$$

drop t_0, redundant

Taylor series in space (RHS)

$$f(x-\Delta x;t|x_0) = f(x;t|x_0) - \Delta x \frac{\partial}{\partial x} f(x;t|x_0) + \frac{1}{2}\Delta x^2 \frac{\partial^2}{\partial x^2} f(x;t|x_0) + \mathcal{O}(\Delta x^3)$$

again, drop t_0

Substitute expansions into the CK equation

$$f(x;t\,|\,x_0) + \Delta t \frac{\partial}{\partial t} f(x;t\,|\,x_0) = \int_{-\infty}^{\infty} \pi_{\Delta x, \Delta t} \Big[f(x;t\,|\,x_0) - \Delta x \frac{\partial}{\partial x} f(x;t\,|\,x_0) + \frac{1}{2}\Delta x^2 \frac{\partial^2}{\partial x^2} f(x;t\,|\,x_0) \Big]\, d(\Delta x)$$

Expand RHS

$$f(x;t\,|\,x_0) + \Delta t \frac{\partial}{\partial t} f(x;t\,|\,x_0) = \int_{-\infty}^{\infty} \pi_{\Delta x, \Delta t}\, f(x;t\,|\,x_0)\, d(\Delta x)$$

$$- \int_{-\infty}^{\infty} \pi_{\Delta x, \Delta t}\, \Delta x \frac{\partial}{\partial x} f(x;t\,|\,x_0)\, d(\Delta x)$$

$$+ \frac{1}{2} \int_{-\infty}^{\infty} \pi_{\Delta x, \Delta t}\, \Delta x^2 \frac{\partial^2}{\partial x^2} f(x;t\,|\,x_0)\, d(\Delta x)$$

Underlined terms do not depend on Δx and can come out of the integral.

$$\int_{-\infty}^{\infty} \pi_{\Delta x, \Delta t}\, f \, d(\Delta x) = f \underbrace{\int_{-\infty}^{\infty} \pi_{\Delta x, \Delta t}\, d(\Delta x)}_{1 \text{ (total probability)}}$$

$$\int_{-\infty}^{\infty} \pi_{\Delta x, \Delta t}\, \Delta x \frac{\partial f}{\partial x}\, d(\Delta x) = \frac{\partial f}{\partial x} \underbrace{\int_{-\infty}^{\infty} \pi_{\Delta x, \Delta t}\, \Delta x \, d(\Delta x)}_{E[\Delta x]}$$

$$\frac{1}{2}\int_{-\infty}^{\infty} \pi_{\Delta x, \Delta t}\, \Delta x^2 \frac{\partial^2 f}{\partial x^2}\, d(\Delta x) = \frac{1}{2} \frac{\partial^2 f}{\partial x^2} \underbrace{\int_{-\infty}^{\infty} \pi_{\Delta x, \Delta t}\, \Delta x^2 \, d(\Delta x)}_{E[\Delta x^2]}$$

Plug into our equation

$$\cancel{f(x;t\,|\,x_0)} + \Delta t \frac{\partial}{\partial t} f(x;t\,|\,x_0) = \cancel{f(x;t\,|\,x_0)} - E[\Delta x] \frac{\partial}{\partial x} f(x;t\,|\,x_0) + \frac{1}{2} E[(\Delta x)^2] \frac{\partial^2}{\partial x^2} f(x;t\,|\,x_0)$$

Divide both sides by Δt and define

$$a = \lim_{\Delta t \to 0} \frac{E[\Delta x]}{\Delta t} \quad \text{Drift function}$$

$$B = \lim_{\Delta t \to 0} \frac{E[(\Delta x)^2]}{\Delta t} \quad \text{Diffusion function}$$

This gives the celebrated Fokker-Planck equation

$$\boxed{\frac{\partial}{\partial t} f(x;t|x_0,t_0) = -a \frac{\partial}{\partial x} f(x;t|x_0,t_0) + \frac{B}{2} \frac{\partial^2}{\partial x^2} f(x;t|x_0;t_0)}$$

(ii) Notes:

- if we keep track of the x-dependence of a and B, we get

$$\frac{\partial f}{\partial t} = -\frac{\partial}{\partial x}(a(x) f) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(B(x) f) \quad \text{Itô calculus version}$$

- why not higher order moments?

  The functions $E(\Delta x)/\Delta t$ and $E(\Delta x^2)/\Delta t$ need to converge in the limit that $\Delta t \to 0$.

  - Drift terms (a) are like convection. you move distance $a\Delta t$ in time Δt.
  - Diffusion terms come from a variance $\propto \Delta t$ (e.g., Wiener process) so you spread out a distance $B\Delta t$ in time Δt.
  - Higher order terms (Δx^n, $n>2$) scale like $\Delta t^{n/2}$, so they go to zero in the limit that $\Delta t \to 0$.

D. Using the Fokker-Planck equation

we can solve the Fokker-Planck equation to find our process PDF! we just need to set a and B (drift, diffusion using our physical intuition) and initial and boundary conditions! If we need, we can always get marginal or joint PDFs from our other familiar relationships.

(i) Example: Wiener Process

let $a=0$, $B=1$, $f=f(x,t|x_0,t_0)$

$$\frac{\partial f}{\partial t}=\frac{1}{2}\frac{\partial^2 f}{\partial x^2}$$

Initial condition

$$f(x,t_0|x_0,t_0)=\delta(x-x_0)$$

Deterministic I.C. At $x=x_0$.

Boundary Condition

$$f(x,t|x_0,t_0)\to 0$$

at $\pm\infty$.

Quick solve w/ Fourier transform.

Fourier Transform PDE: $$\frac{\partial \hat{f}}{\partial t}=-\frac{1}{2}k^2\hat{f}$$

Solve ODE in Fourier space: $\hat{f}=C\exp(-\frac{1}{2}k^2t)$

Take FT of I.C.: $\hat{f}(t=t_0)=\exp(ikx_0)$

Find integration constant: $\hat{f}(t=t_0)=C\exp(-\frac{1}{2}k^2t_0)=\exp(ikx_0)$

$$C=\exp(ikx_0+\tfrac{1}{2}k^2t_0)$$

Fourier Transform Solution:

$$\hat{f}(t)=\exp(ikx_0-\tfrac{1}{2}k^2(t-t_0))$$

Take Inverse FT:

$$f(x,t|x_0,t_0)=\frac{1}{\sqrt{2\pi(t-t_0)}}\exp\left(-\frac{1}{2}\frac{(x-x_0)^2}{(t-t_0)}\right)$$

E. Stationary Processes

A random process is strictly stationary of order n if the joint probability density does not depend on Δt.

$$f[x(t_1),x(t_2),\dots,x(t_n)]=$$
$$f[x(t_1+\Delta t),x(t_2+\Delta t),\dots,x(t_n+\Delta t)]$$

(The nth joint pdf is the same as you translate in time).

A random process is wide-sense stationary if

$$m(t) = \text{constant}$$

$$R(t_1, t_2) = R(t_1 - t_2) \text{ only.}$$

strict implies wide, but inverse is not guaranteed.

When a markov process has a constant (i.e. homogeneous) transition probability $f(x_i \mid x_{i-1})$, then (under some mild conditions*) the process will converge at long times to a stationary distribution.

* The system must be ergodic, meaning the process will eventually visit all possible states of the R.V. (no forbidden states).

Examples

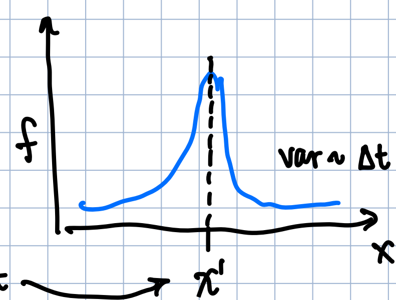
- white noise is a stationary process (but not Markovian, independent).
- The Wiener process is not stationary
- The O-U process is not stationary, but it becomes stationary in the limit that $t \to \infty$.
- The steady state solution of the Fokker-Planck equation is the stationary distribution of a stochastic process.

F. Appendix: Short-time Wiener kernel

Weiner process kernel

$$f(x; t \mid x'; s) = \frac{1}{\sqrt{2\pi(t-s)}} \exp\left[-\frac{1}{2} \frac{(x-x')^2}{t-s}\right]$$

$$f(x; t+\Delta t \mid x-\Delta x; t) = \frac{1}{\sqrt{2\pi\Delta t}} \exp\left[-\frac{1}{2} \frac{\Delta x^2}{\Delta t}\right]$$



where it was before, $x \mid x'$.
So, centered at old point