

Math 106 – Notes

Week 3: January 22, 2026

Ethan Levien

Motivation

We've seen that we can view a stochastic process as a random variable on the space of sample paths $\Omega = \Omega_0^{\mathbf{T}}$ where in this class $\mathbf{T} = \mathbb{R}$ or $\mathbf{T} = \mathbb{N}$.¹ The Kolmogorov Extension Theorem tells us that if all the finite dimensional distributions are defined, then there is some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ which is consistent with these finite dimensional distributions. Although we don't usually work directly with the entirety of \mathcal{F} , it is important to understand which events in \mathcal{F} we can ask about (that is, evaluate probabilities of), based on observations of a process up to a given time. This motivates the idea of a filtration.

Filtrations

Definition 1 (Filtration and adapted). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space associated with a stochastic process.*

- (D5.3) A filtration $\{\mathcal{F}_t\}_{t \geq 0}$ is a family of σ -algebras on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$ with $0 \leq s < t$.
- A process $\{X_t\}$ is said to be \mathcal{F}_t -adapted if X_t is \mathcal{F}_t measurable. That is, as a function $X_t(\cdot) : \Omega \rightarrow \mathbb{R}$, $X_t^{-1}(B) \in \mathcal{F}_t$ for $B \in \mathcal{R}$.
- We use \mathcal{F}_t^X to denote the smallest filtration such that the process $\{X_t\}$ is \mathcal{F}_t -adapted.

In words, \mathcal{F}_t is simply all the events whose probabilities can be determined by looking at realizations of a process up to time t . Importantly, these events are subsets of the larger σ -algebra \mathcal{F} , which is important to point out when we compare the following example to that in the book.

Examples

Two state discrete time Markov Chain

Consider a two-state discrete-time Markov chain; hence we can set $\Omega_0 = \{0, 1\}$ and Ω is the space of all infinite binary sequences $\omega = (\omega_0, \omega_1, \omega_2, \dots)$ with $\omega_i \in \Omega_0$. To be concrete, let's flip a fair coin to get the first value: $\mathbb{P}(X_0 = 0) = 1/2$. Remember that we can think of X_n as the *coordinate process*, which when evaluated on the entire sequence returns the n th value: $X_n(\omega) = \omega_n$.

¹I use Ω_0 to denote the state space of the process. In the book they use S or Ω , but this is confusing since Ω is also used for the path space in this chapter.

Consider the filtration \mathcal{F}_t^X for this process. At $n = 0$ we only know the first element in the sequence, thus

$$\mathcal{F}_0^X = \{\emptyset, \Omega, \{\omega \in \Omega : \omega_0 = 0\}, \{\omega \in \Omega : \omega_0 = 1\}\} = \{\emptyset, \Omega, C(0), C(1)\}. \quad (1)$$

Looking at each element in \mathcal{F}_0^X , we should be able to evaluate its probability based on only the observation of X_0 , and this is indeed the case: the probabilities are 0, 1, 1/2, and 1/2. It is useful to introduce the idea and notation of *cylinder sets*, which are the sets

$$C(\varepsilon_0, \dots, \varepsilon_n) = \{\omega \in \Omega : \omega_i = \varepsilon_i, 0 \leq i \leq n\}. \quad (2)$$

In words, $C(\varepsilon_0, \dots, \varepsilon_n)$ is the set of all sequences in Ω for which the first n elements are given, and the given values are $\varepsilon_0, \dots, \varepsilon_n$. Hence $\mathcal{F}_0^X = \{\emptyset, \Omega, C(0), C(1)\}$. We can also write this as $\mathcal{F}_0^X = \sigma(C(0), C(1))$ which means the σ -algebra generated by the two cylinder sets $C(0)$ and $C(1)$. The other sets, \emptyset and Ω are obtained by $\Omega = C(0) \cup C(1)$ and $\emptyset = (C(0) \cup C(1))^c$.

Remark 1. Note that technically I could have written $\mathcal{F}_0^X = \sigma(C(0))$ since in this case $C(1) = C(0)^c$, but it is conceptually helpful to keep track of all cylinder sets of a given order since this will generalize to different state spaces more naturally.

Similarly, at time 1 we know the first two coordinates. The atoms of the σ -algebra are

$$C(0,0), C(0,1), C(1,0), C(1,1).$$

Thus

$$\mathcal{F}_1^X = \sigma(C(0,0), C(0,1), C(1,0), C(1,1)), \quad (3)$$

that is, all unions of these four cylinder sets. You can check that $\mathcal{F}_0^X \subset \mathcal{F}_1^X$.

Remark 2. Note that in the textbook, on page 105, they use the notation $\{H \cdot\}$, for example, to denote the cylinder set $C(H)$. But importantly, they don't always write the \cdot and hence the notation for \mathcal{F}_2^X appears to include sets that are not actually subsets of Ω , such as $\{H\}$. Note that in that example use $n = 1$ as the first value in the sequence, but for stochastic process (elsewhere in the book), we have been using $n = 0$.

One-jump Q -process

Consider the Q -process on $\{0, 1\}$ with Q matrix

$$Q = \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix} \quad (4)$$

and assume that $X_0 = 0$. This is a process that has only one jump time J_1 at which X_t jumps from 0 to 1. Remember that for a general continuous time stochastic process we have a rather complicated path space $\Omega^{\mathbb{R}}$. However, in this case we can identify the underlying space as the much simpler $\Omega = [0, \infty)$ by letting the element $\omega \in \Omega$ represent the (random) jump time $J_1(\omega) = \omega$.

Here, the scalar value ω is playing the role that the entire sequence played in the previous example, and we can get away with this precisely because the trajectory can be constructed based only on this value. In particular, the function $X_t(\cdot) : \Omega \rightarrow \Omega_0 \subset \mathbb{R}$ is given by

$$X_t(\omega) = 1_{\{t < \omega\}} = \begin{cases} 0, & t < \omega, \\ 1, & t \geq \omega. \end{cases} \quad (5)$$

This induces a very simple filtration, which can be written as

$$\mathcal{F}_t^X = \sigma(\{J_1 \leq s\} : 0 \leq s \leq t). \quad (6)$$

Events such as $\{J_1 \in (t_1, t_2)\}$ for $0 \leq t_1 < t_2 \leq t$ are therefore in \mathcal{F}_t^X .

General structure of filtrations for Q -processes

For a general Q -process on $\Omega_0 = \{1, \dots, I\}$ we can also describe the filtration rather simply. Between jump times the process is constant, so a path is determined by a sequence

$$((Y_0, Y_1, Y_2, \dots), (J_1, J_2, J_3, \dots)),$$

with the convention that if the process eventually becomes absorbing then the remaining jump times are $+\infty$. We can therefore identify Ω as with the space

$$\Omega = \Omega_0^{\mathbb{N} \cup \{0\}} \times \mathbb{R}^{\mathbb{N}}. \quad (7)$$

The filtration of a general Q -process can be generated by

$$\mathcal{F}_t^X = \sigma\left(\{Y_0 = i\}, \{Y_n = i\} \cap \{J_n \leq s\} : 0 \leq s \leq t, n \geq 0, i \in \Omega_0\right). \quad (8)$$

In words, \mathcal{F}_t^X contains exactly the information about:

- The initial value of the chain
- which jumps have occurred by time $s \leq t$,
- the values of the embedded chain for those jumps that have occurred (which includes the current state of the process).

Remark 3. *This decomposition illustrates the strong Markov property (mentioned after T3.15): the future evolution after time t depends only on the current state X_t and not on the past jump structure, even though the past jump times are fully encoded in the filtration. It also clarifies why the Gillespie algorithm can generate waiting times and next states separately: the holding time in a state and the next jump of the embedded chain are conditionally independent given the current state, and both pieces of information appear naturally in \mathcal{F}_t^X .*

Note that other events such as $\{J_n \leq s\}$ can be written as unions of these:

$$\{J_n \leq s\} = \bigcup_{i \in \Omega_0} \{Y_n = i\} \cap \{J_n \leq s\}, \quad (9)$$

However, the events $\{Y_n = i\}$ are not in \mathcal{F}_t^X since even if we take the union over all events with $\{Y_n = i\}$, we are left with $\{Y_n = i\} \cap \{J_n \leq t\}$. We would need to take $t \rightarrow \infty$ to obtain $\{Y_n = i\}$.

Stopping times

Stopping times formalize the idea of a random time whose occurrence can be detected by observing a process up to that time. Let $\{\mathcal{F}_t\}_{t \geq 0}$ be a filtration on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 2 (Stopping time D5.4). *A random time $T : \Omega \rightarrow [0, \infty]$ is called a stopping time with respect to the filtration \mathcal{F}_t if for every $t \geq 0$,*

$$\{T \leq t\} \in \mathcal{F}_t. \quad (10)$$

This definition works in both discrete and continuous time. Note that just like X_t , T is viewed here as a function from Ω to (as subset of) \mathbb{R} . Note that, taking the function view of T , we can write

$$\{T \leq t\} = T^{-1}([0, t]) \quad (11)$$

Whether or not T has occurred by time t must be determined entirely from the information available in \mathcal{F}_t . A key point is that stopping times are defined through the sets $\{T \leq t\}$ rather than the sets $\{T = t\}$ (see Exercise 5.1). The latter do not behave well in continuous time and in particular, they generally have probability zero. For example, in a Q -process $\mathbb{P}(J_1 = t) = 0$ for every fixed t . The definition of a stopping time uses $\{T \leq t\}$ —not because $\{T = t\}$ is unobservable, but because $\{T = t\}$ is useless (measure zero, no monotonicity, no functional measurability). The event $\{T \leq t\}$ is the one that makes T a measurable, well-behaved random time.

Combining stopping times

If we have multiple stopping times, they generally behave well if we take their min, max, inf, sup. This is what P5.5 is saying. Let's see how this works out for the max. If T_1 and T_2 are stopping times, let $T = T_1 \vee T_2$ (note \vee mean max and \wedge means min).

$$\{T \leq t\} = \{T_1 \leq t\} \cup \{T_2 \leq t\} \in \mathcal{F}_t^X. \quad (12)$$

Intuitively it should make sense: We can determine if the larger time has happened yet by observing the processes up to time T . You should prove the rest of P5.5 (which is exercises 5.2) on your own.

Other examples

First time to see a sequence

We now give a simple example of a stopping time in the discrete-time setting using the cylinder sets introduced above for a binary process. Fix a binary word of length $m + 1$,

$$(\varepsilon_0, \dots, \varepsilon_m) \in \{0, 1\}^{m+1}.$$

We define the first time this word appears as a consecutive block in the sequence $X = (X_0, X_1, \dots)$ by

$$T = \inf \{n \geq m : (X_{n-m}, \dots, X_n) = (\varepsilon_0, \dots, \varepsilon_m)\}.$$

We claim that T is a stopping time with respect to the filtration $\{\mathcal{F}_n^X\}_{n \geq 0}$. For a fixed integer k with $m \leq k \leq n$, consider the event

$$\{(X_{k-m}, \dots, X_k) = (\varepsilon_0, \dots, \varepsilon_m)\}.$$

This means that the coordinates X_{k-m}, \dots, X_k are fixed and equal to $\varepsilon_0, \dots, \varepsilon_m$, while the earlier coordinates X_0, \dots, X_{k-m-1} may take either value 0 or 1. Therefore the event can be written explicitly as the finite union

$$\{(X_{k-m}, \dots, X_k) = (\varepsilon_0, \dots, \varepsilon_m)\} = \bigcup_{(\eta_0, \dots, \eta_{k-m-1}) \in \{0, 1\}^{k-m}} C(\eta_0, \dots, \eta_{k-m-1}, \varepsilon_0, \dots, \varepsilon_m),$$

where each term in the union is the cylinder set specifying the first $k+1$ coordinates of the sequence.

Holding times and the difference of stopping times

Consider a Q -process and recall that H_n denotes the holding time. Is H_n a stopping time? First think about this intuitively. If we observe the process to time t can we determine if $H_n \leq t$? The answer is no since if the n th jump happens after t then we can't determine H_n . This is true even though H_n is the difference of stopping times. Consider that

$$\{H_n \leq t\} = \{J_n - J_{n-1} \leq t\} = \{J_n \leq t + J_{n-1}\} \in \mathcal{F}_{t+J_{n-1}}^X \quad (13)$$

This $\{H_n \leq t\}$ is determined by events $+J_{n-1}$ in the future. In general, the difference of stopping times are not stopping times.