Chapter 6

Random Times and Their Properties

Section 6.1 recalls the definition of a filtration (a growing collection of σ-fields) and of "stopping times" (basically, measurable random times).

Section 6.2 defines various sort of "waiting" times, including hitting, first-passage, and return or recurrence times.

Section 6.3 proves the Kac recurrence theorem, which relates the finite-dimensional distributions of a stationary process to its mean recurrence times.

6.1 Reminders about Filtrations and Stopping Times

You will have seen these in 36-752 as part of martingale theory, though their application is more general, as we'll see.

Definition 55 (Filtration) *Let T be an ordered index set. A collection $\mathcal{F}_t$, $t \in T$ of σ-algebras is a* filtration *(with respect to this order) if it is non-decreasing, i.e., $f \in \mathcal{F}_t$ implies $f \in \mathcal{F}_s$ for all $s > t$. We generally abbreviate this filtration by $\{\mathcal{F}\}_t$. Define $\{\mathcal{F}\}_{t^+}$ as $\bigcap_{s>t} \mathcal{F}_s$. If $\{\mathcal{F}\}_{t^+} = \{\mathcal{F}\}_t$, then $\{\mathcal{F}\}_t$ is* right-continuous.

Recall that we generally think of a σ-algebra as representing available information — for any event $f \in \mathcal{F}$, we can answer the question "did f happen?" A filtration is a way of representing our information about a system growing over time. To see what right-continuity is about, imagine it failed, which would mean $\mathcal{F}_t \subset \bigcap_{s>t} \mathcal{F}_s$. Then there would have to be events which were detectable at *all* times after t, but not at t itself, i.e., some sudden jump in our information right after t. This is what right-continuity rules out.

Definition 56 (Adapted Process) *A stochastic process X on T is* adapted *to a filtration $\{\mathcal{F}\}_t$ if $\forall t$, X_t is $\mathcal{F}_t$-measurable. Any process is adapted to the filtration it induces, $\sigma\{X_s : \ s \leq t\}$. This* natural filtration *is written $\left\{\mathcal{F}^X\right\}_t$.*

A process being adapted to a filtration just means that, at every time, the filtration gives us enough information to find the value of the process.

Definition 57 (Stopping Time, Optional Time) *An* optional time *or a* stopping time*, with respect to a filtration $\{\mathcal{F}\}_t$, is a T-valued random variable τ such that, for all t,*

$$\{\omega \in \Omega : \ \tau(\omega) \leq t\} \quad \in \quad \mathcal{F}_t \tag{6.1}$$

If Eq. 6.1 holds with $<$ instead of $\leq$, then τ is weakly optional *or a* weak stopping time*.*

Basically, all we're doing here is defining what we mean by "a random time at which something detectable happens". That being the case, it is natural to ask what information we have when that detectable thing happens.

Definition 58 ($\mathcal{F}_\tau$ for a Stopping Time τ) *If τ is a $\{\mathcal{F}\}_t$ stopping time, then the σ-algebra $\mathcal{F}_\tau$ is given by*

$$\mathcal{F}_\tau \quad \equiv \quad \{A \in \mathcal{F} : \forall t, \ A \cap \{\omega : \ \tau(\omega) \leq t\} \in \mathcal{F}_t\} \tag{6.2}$$

I admit that the definition of $\mathcal{F}_\tau$ looks bizarre, and I won't blame you if you have to read it a few times to convince yourself it isn't circular. Here is a simple case where it makes sense. Let X be a one-sided process, and τ a discrete $\left\{\mathcal{F}^X\right\}_t$ stopping time. Then

$$\mathcal{F}^X_\tau = \sigma\left(X(t \wedge \tau) : \ t \geq 0\right) \tag{6.3}$$

That is, $\mathcal{F}^X_\tau$ is everything we know from observing X up to time τ. (This is Exercise 6.2.) The convoluted-looking definition of $\mathcal{F}_\tau$ carries this idea over to the more general situation where τ is continuous and we don't necessarily have a single variable generating the filtration. A filtration lets us tell whether some event A happened by the random time τ if simultaneously gives us enough information to notice τ and A.

The process $Y(t) = X(t \wedge \tau)$ is follows along with X up until τ, at which point it becomes fixed in place. It is accordingly called an *arrested*, *halted* or *stopped* version of the process. This seems to be the origin of the name "stopping time".

6.2 Waiting Times

"Waiting times" are particular kinds of optional kinds: how much time must elapse before a given event happens, either from a particular starting point, or averaging over all trajectories? Often, these are of particular interest in themselves, and some of them can be related to other quantities of interest.

Definition 59 (Hitting Time) *Given a one-sided Ξ-valued process X, the hitting time τ_B of a measurable set $B \subset \Xi$ is the first time at which $X(t) \in B$;*

$$\tau_B = \inf \{t > 0 : X_t \in B\} \tag{6.4}$$

Example 60 (Fixation through Genetic Drift) *Consider the variation in a given locus (roughly, gene) in an evolving population. If there are k different versions of the gene ("alleles"), the state of the population can be represented by a vector $X(t) \in \mathbb{R}^k$, where at each time $X_i(t) \geq 0$ and $\sum_i X_i(t) = 1$. This set is known as the k-dimensional* probability simplex *S_k. We say that a certain allele has* been fixed *in the population or* gone to fixation *at t if $X_i(t) = 1$ for some i, meaning that all members of the population have that version of the gene. Fixation corresponds to $X(t) \in V$, where V consists of the vertices of the simplex. An important question in evolutionary theory is how long it takes the population to go to fixation. By comparing the actual rate of fixation to that expected under a model of adaptively-neutral genetic drift, it is possible to establish that some genes are under the influence of natural selection.*

Gillespie (1998) is a nice introduction to population genetics, including this problem among many others, using only elementary probability. More sophisticated models treat populations as measure-valued stochastic processes.

Example 61 (Stock Options) *A* stock option[1] *is a legal instrument giving the holder the right to buy a stock at a specified price (the* strike price*, c) before a certain expiration date t_e. The point of the option is that, if you exercise it at a time t when the price of the stock $p(t)$ is above c, you can turn around and sell the stock to someone else, making a profit of $p(t) - c$. When $p(t) > c$, the option is said to be* in money *or* above water*. Options can themselves be sold, and the value of an option depends on how much money it could be used to make, which in turn depends on the probability that it will be "in money" before time t_e. An important part of mathematical finance thus consists of problems of the form "assuming prices $p(t)$ follow a process distribution μ, what is the distribution of hitting times of the set $p(t) > c$?"*

While the financial industry is a major consumer of stochastics, and it has a legitimate role to play in capitalist society, I do hope you will find something more interesting to do with your new-found mastery of random processes, so I will not give many examples of this sort. If you want much, much more, read Shiryaev (1999).

Definition 62 (First Passage Time) *When $\Xi = \mathbb{R}$ or $\mathbb{Z}$, we call the hitting time of the origin the time of* first passage through the origin*, and similarly for other points.*

[1] Actually, this is just one variety of option (an "American call"), out of a huge variety. I will *not* go into details.

Definition 63 (Return Time, Recurrence Time) *Fix a set $B \in \Xi$. Suppose that $X(t_0) \in B$. Then the* return time *or* first return time *of B is* recurrence time *of B is* $\inf \{t > t_0 \ : \ X(t) \in B\}$*, and the* recurrence time θ_B *is the difference between the first return time and t_0.*

Note 1: If I'm to be honest with you, I should admit that "return time" and "recurrence time" are used more or less interchangeably in the literature to refer to either the time *coordinate* of the first return (what I'm calling the return time) or the time *interval* which elapses before that return (what I'm calling the recurrence time). I will try to keep these straight here. Check definitions carefully when reading papers!

Note 2: Observe that if we have a discrete-parameter process, and are interested in recurrences of a finite-length sequence of observations $w \in \Xi^k$, we can handle this situation by the device of working with the shift operator in sequence space.

The question of whether any of these waiting times is optional (i.e., measurable) must, sadly, be raised. The following result is generally enough for our purposes.

Proposition 64 (Some Sufficient Conditions for Waiting Times to be Weakly Optional) *Let X be a Ξ-valued process on a one-sided parameter T, adapted to a filtration $\{\mathcal{F}\}_t$, and let B be an arbitrary measurable set in Ξ. Then τ_B is weakly $\{\mathcal{F}\}_t$-optional under any of the following (sufficient) conditions, and $\{\mathcal{F}\}_t$-optional under the first two:*

1. *T is discrete.*
2. *T is $\mathbb{R}^+$, Ξ is a metric space, B is closed, and $X(t)$ is a continuous function of t.*
3. *T is $\mathbb{R}^+$, Ξ is a topological space, B is open, and $X(t)$ is right-continuous as a function of t.*

Proof: See, for instance, Kallenberg, Lemma 7.6, p. 123. □

6.3 Kac's Recurrence Theorem

For strictly stationary, discrete-parameter sequences, a very pretty theorem, due to Mark Kac (1947), relates the probability of seeing a particular event to the mean time between recurrences of the event. Throughout, we consider an arbitrary Ξ-valued process X, subject only to the requirements of stationarity and a discrete parameter.

Fix an arbitrary measurable set $A \in \Xi$ with $\mathbb{P}(X_1 \in A) > 0$, and consider a new process $Y(t)$, where $Y_t = 1$ if $X_t \in A$ and $Y_t = 0$ otherwise. By Exercise 5.1, Y_t is also stationary. Thus $\mathbb{P}(X_1 \in A, X_2 \notin A) = \mathbb{P}(Y_1 = 1, Y_2 = 0)$. Let us abbreviate $\mathbb{P}(Y_1 = 0, Y_2 = 0, \ldots Y_{n_1} = 0, Y_n = 0)$ as w_n; this is the probability of making n consecutive observations, none of which belong to the event

A. Clearly, $w_n \geq w_{n+1}$. Similarly, let $e_n = \mathbb{P}(Y_1 = 1, Y_2 = 0, \ldots Y_n = 0)$ and $r_n = \mathbb{P}(Y_1 = 1, Y_2 = 0, \ldots Y_n = 1)$ — these are, respectively, the probabilities of starting in A and not returning within $n-1$ steps, and of starting in A and returning for the first time after $n-2$ steps. (Set e_1 to $\mathbb{P}(Y_1 = 1)$, and $w_0 = e_0 = 1$.)

Lemma 65 (Some Recurrence Relations for Kac's Theorem) *The following recurrence relations hold among the probabilities w_n, e_n and r_n:*

$$\begin{aligned} e_n &= w_{n-1} - w_n, \ n \geq 1 && (6.5) \\ r_n &= e_{n-1} - e_n, \ n \geq 2 && (6.6) \\ r_n &= w_{n-2} - 2w_{n-1} + w_n, \ n \geq 2 && (6.7) \end{aligned}$$

Proof: To see the first equality, notice that

$$\begin{aligned} &\mathbb{P}(Y_1 = 0, Y_2 = 0, \ldots Y_{n-1} = 0) && (6.8) \\ &= \mathbb{P}(Y_2 = 0, Y_3 = 0, \ldots Y_n = 0) \\ &= \mathbb{P}(Y_1 = 1, Y_2 = 0, \ldots Y_n = 0) + \mathbb{P}(Y_1 = 0, Y_2 = 0, \ldots Y_n = 0) && (6.9) \end{aligned}$$

using first stationarity and then total probability. To see the second equality, notice that, by total probability,

$$\begin{aligned} &\mathbb{P}(Y_1 = 1, Y_2 = 0, \ldots Y_{n-1} = 0) && (6.10) \\ &= \mathbb{P}(Y_1 = 1, Y_2 = 0, \ldots Y_{n-1} = 0, Y_n = 0) + \mathbb{P}(Y_1 = 1, Y_2 = 0, \ldots Y_{n-1} = 0, Y_n = 1) \end{aligned}$$

The third relationship follows from the first two. □

Theorem 66 (Recurrence in Stationary Processes) *Let X be a Ξ-valued discrete-parameter stationary process. For any set A with $\mathbb{P}(X_1 \in A) > 0$, for almost all ω such that $X_1(\omega) \in A$, there exists a τ for which $X_\tau(\omega) \in A$.*

$$\sum_{k=1}^{\infty} \mathbb{P}(\theta_A = k | X_1 \in A) = 1 \quad (6.11)$$

Proof: The event $\{\theta_A = k, X_1 \in A\}$ is the same as the event $\{Y_1 = 1, Y_2 = 0, \ldots Y_{k+1} = 1\}$. Since $\mathbb{P}(X_1 \in A) > 0$, we can handle the conditional probabilities in an elementary fashion:

$$\begin{aligned} \mathbb{P}(\theta_A = k | X_1 \in A) &= \frac{\mathbb{P}(\theta_A = k, X_1 \in A)}{\mathbb{P}(X_1 \in A)} && (6.12) \\ &= \frac{\mathbb{P}(Y_1 = 1, Y_2 = 0, \ldots Y_{k+1} = 1)}{\mathbb{P}(Y_1 = 1)} && (6.13) \\ \sum_{k=1}^{\infty} \mathbb{P}(\theta_A = k | X_1 \in A) &= \frac{\sum_{k=1}^{\infty} \mathbb{P}(Y_1 = 1, Y_2 = 0, \ldots Y_{k+1} = 1)}{\mathbb{P}(Y_1 = 1)} && (6.14) \\ &= \frac{\sum_{k=2}^{\infty} r_k}{e_1} && (6.15) \end{aligned}$$

Now consider the finite sums, and apply Eq. 6.7.

$$\sum_{k=2}^{n} r_k = \sum_{k=2}^{n} w_{k-2} - 2w_{k-1} + w_k \quad (6.16)$$

$$= \sum_{k=0}^{n-2} w_k + \sum_{k=2}^{n} w_k - 2\sum_{k=1}^{n-1} w_k \quad (6.17)$$

$$= w_0 + w_n - w_1 - w_{n-1} \quad (6.18)$$

$$= (w_0 - w_1) - (w_{n-1} - w_n) \quad (6.19)$$

$$= e_1 - (w_{n-1} - w_n) \quad (6.20)$$

where the last line uses Eq. 6.6. Since $w_{n-1} \geq w_n$, there exists a $\lim_n w_n$, which is ≥ 0 since every individual w_n is. Hence $\lim_n w_{n-1} - w_n = 0$.

$$\sum_{k=1}^{\infty} \mathbb{P}(\theta_A = k | X_1 \in A) = \frac{\sum_{k=2}^{\infty} r_k}{e_1} \quad (6.21)$$

$$= \lim_{n \to \infty} \frac{e_1 - (w_{n-1} - w_n)}{e_1} \quad (6.22)$$

$$= \frac{e_1}{e_1} \quad (6.23)$$

$$= 1 \quad (6.24)$$

which was to be shown. □

Corollary 67 (Poincaré Recurrence Theorem) *Let F be a transformation which preserves measure μ. Then for any set A of positive μ measure, for μ-almost-all $x \in A$, $\exists n \geq 1$ such that $F^n(x) \in A$.*

Proof: A direct application of the theorem, given the relationship between stationary processes and measure-preserving transformations we established by Corollary 54. □

Corollary 68 ("Nietzsche") *In the set-up of the previous theorem, if $X_1(\omega) \in A$, then $X_t \in A$ for infinitely many t (a.s.).*

Proof: Repeated application of the theorem yields an infinite sequence of times $\tau_1, \tau_2, \tau_3, \ldots$ such that $X_{\tau_i}(\omega) \in A$, for almost all ω such that $X_1(\omega) \in A$ in the first place. □

Now that we've established that once something happens, it will happen again and again, we would like to know how long we have to wait between recurrences.

Theorem 69 (Kac's Recurrence Theorem) *Continuing the previous notation, $\mathbf{E}[\theta_A | X_1 \in A] = 1/\mathbb{P}(X_1 \in A)$ if and only if $\lim_n w_n = 0$.*

PROOF: "If": Unpack the expectation:

$$\mathbf{E}\left[\theta_A | X_1 \in A\right] = \sum_{k=1}^{\infty} k \frac{\mathbb{P}\left(Y_1 = 1, Y_2 = 0, \ldots Y_{k+1} = 1\right)}{\mathbb{P}\left(Y_1 = 1\right)} \quad (6.25)$$

$$= \frac{1}{\mathbb{P}\left(X_1 \in A\right)} \sum_{k=1}^{\infty} k r_{k+1} \quad (6.26)$$

so we just need to show that the last series above sums to 1. Using Eq. 6.7 again,

$$\sum_{k=1}^{n} k r_{k+1} = \sum_{k=1}^{n} k(w_{k-1} - 2w_k + w_{k+1}) \quad (6.27)$$

$$= \sum_{k=1}^{n} k w_{k-1} + \sum_{k=1}^{n} k w_{k+1} - 2\sum_{k=1}^{n} k w_k \quad (6.28)$$

$$= \sum_{k=0}^{n-1} (k+1) w_k + \sum_{k=2}^{n+1} (k-1) w_k - 2\sum_{k=1}^{n} k w_k \quad (6.29)$$

$$= w_0 + n w_{n+1} - (n+1) w_n \quad (6.30)$$

$$= 1 - w_n - n(w_n - w_{n+1}) \quad (6.31)$$

We therefore wish to show that $\lim_n w_n = 0$ implies $\lim_n w_n + n(w_n - w_{n+1}) = 0$. By hypothesis, it is enough to show that $\lim_n n(w_n - w_{n+1}) = 0$. The partial sums on the left-hand side of Eq. 6.27 are non-decreasing, so $w_n + n(w_n - w_{n+1})$ is non-increasing. Since it is also ≥ 0, the limit $\lim_n w_n + n(w_n - w_{n+1})$ exists. Since $w_n \to 0$, $w_n - w_{n+1} \leq w_n$ must also go to zero; the only question is whether it goes to zero fast enough. So consider

$$\lim_n \sum_{k=1}^{n} w_k - \sum_{k=1}^{n} w_{k+1} \quad (6.32)$$

Telescoping the sums again, this is $\lim_n w_1 - w_{n+1}$. Since $\lim_{n+1} w_{n+1} = \lim_n w_n = 0$, the limit exists. But we can equally re-write Eq. 6.32 as

$$\lim_n \sum_{k=1}^{n} w_k - w_{k+1} = \sum_{n=1}^{\infty} w_n - w_{n+1} \quad (6.33)$$

Since the sum converges, the individual terms $w_n - w_{n+1}$ must be $o(n^{-1})$. Hence $\lim_n n(w_n - w_{n+1}) = 0$, as was to be shown.

"Only if": From Eq. 6.31 in the "if" part, we see that the hypothesis is equivalent to

$$1 = \lim_n \left(1 - w_n - n(w_n - w_{n+1})\right) \quad (6.34)$$

Since $w_n \geq w_{n+1}$, $1 - w_n - n(w_n - w_{n+1}) \leq 1 - w_n$. We know from the proof of Theorem 66 that $\lim_n w_n$ exists, whether or not it is zero. If it is not zero, then $\lim_n \left(1 - w_n - n(w_n - w_{n+1})\right) \leq 1 - \lim_n w_n < 1$. Hence $w_n \to 0$ is a necessary condition. □

Example 70 (Counter-example for Kac's Recurrence Theorem) *One might imagine that the condition $w_n \to 0$ in Kac's Theorem is redundant, given the assumption of stationarity. Here is a counter-example. Consider a homogeneous Markov chain on a finite space Ξ, which is partitioned into two non-communicating components, Ξ_1 and Ξ_2. Each component is, internally, irreducible and aperiodic, so there will be an invariant measure μ_1 supported on Ξ_1, and another invariant measure μ_2 supported on Ξ_2. But then, for any $s \in [0,1]$, $s\mu_1 + (1-s)\mu_2$ will also be invariant. (Why?) Picking $A \subset \Xi_2$ gives $\lim_n w_n = s$, the probability that the chain begins in the wrong component to ever reach A.*

Kac's Theorem turns out to be the foundation for a fascinating class of methods for learning the distributions of stationary processes, and for "universal" prediction and data compression. There is also an interesting interaction with large deviations theory. This subject is one possibility for further discussion at the end of the course. Whether or not we get there, let me recommend some papers in a footnote.[2]

6.4 Exercises

Exercise 6.1 (Weakly Optional Times and Right-Continuous Filtrations) *Show that a random time τ is weakly $\{\mathcal{F}\}_t$-optional iff it is $\{\mathcal{F}\}_{t^+}$-optional.*

Exercise 6.2 (Discrete Stopping Times and Their σ-Algebras) *Prove Eq. 6.3. Does the corresponding statement hold for two-sided processes?*

Exercise 6.3 (Kac's Theorem for the Logistic Map) *First, do Exercise 5.3. Then, using the same code, suitably modified, numerically check Kac's Theorem for the logistic map with $a = 4$. Pick any interval $I \subset [0,1]$ you like, but be sure not to make it* too *small.*

1. *Generate n initial points in I, according to the invariant measure $\frac{1}{\pi\sqrt{x(1-x)}}$. For each point x_i, find the first t such that $F^t(x_i) \in I$, and take the mean over the sample. What happens to this* space average *as n grows?*

2. *Generate a single point x_0 in I, according to the invariant measure. Iterate it T times. Record the successive times $t_1, t_2, \ldots$ at which $F^t(x_0) \in I$, and find the mean of $t_i - t_{i-1}$ (taking $t_0 = 0$). What happens to this* time average *as T grows?*

[2] Kontoyiannis *et al.* (1998); (Ornstein and Weiss, 1990); Algoet (1992).