CSCI-6971 Lecture Notes: Stochastic processes*

Kristopher R. Beevers
Department of Computer Science
Rensselaer Polytechnic Institute
beevek@cs.rpi.edu

February 2, 2006

1 Overview

Definition 1.1. A *stochastic process* is a collection of random variables $X_1, X_2, \ldots, X_t$ parameterized by time.

There are a number of aspects of a stochastic process that we can examine. Among them:

- dependencies between variables in the sequence
- various kinds of long-term averages
- the frequency at which "boundary events" occur

Some examples of stochastic processes include the following:

Definition 1.2. *Random walks* can be thought of as "cumulative" stochastic processes. For example, $S_n = \sum_{r=1}^{n} X_r + S_0$ is a random walk on the line.

Definition 1.3. *Arrival processes* such as the Bernoulli and Poisson processes model the frequency of "arrivals" or "successes" in some time-dependent model. In general, we consider models in which interarrival times are independent random variables.

Definition 1.4. *Markov processes* are sequences where the next value depends on the past only through the current value, i.e.:

$$P\left(X_{t+1} = k | X_{1:t}\right) = P\left(X_{t+1} = k | X_t\right) \tag{1}$$

Definition 1.5. *Martingales* are processes where the expectation of the next value is exactly the same as the current value, i.e.:

$$\mathbf{E}\left[X_{t+1} | X_{1:t}\right] = X_t \tag{2}$$

Martingales frequently arise in gambling, for example, and they can be thought of as a model of a truly fair game.

*The primary sources for most of this material are: "Introduction to Probability," D.P. Bertsekas and J.N. Tsitsiklis, Athena Scientific, Belmont, MA, 2002; and "Randomized Algorithms," R. Motwani and P. Raghavan, Cambridge University Press, Cambridge, UK, 1995; and "Stochastic Processes & Models," D. Stirzaker, Oxford University Press, New York, 2005; and the author's own notes.

2 Arrival processes

We now discuss two arrival processes, the Bernoulli process and the Poisson process, which are discrete- and continuous-time analogs of each other, respectively.

2.1 Bernoulli process

The Bernoulli process is an arrival process consisting of a sequence of i.i.d. Bernoulli trails[1], each of which takes unit time. The key property of a Bernoulli process is that it is *memoryless*, i.e.:

$$P(T - n = t | T > n) = P(T = t) \tag{3}$$

(This formulation is derived from the definition of conditional probability.)

There are several interesting PDFs that we can examine for a Bernoulli process:

- The number S of arrivals in n trials:

$$f_S(k) = \text{BINOMIAL}[n, p] = \binom{n}{k} p^k (1-p)^{n-k} \tag{4}$$

$$\mathbf{E}[S] = np \tag{5}$$

$$\text{var}(S) = np(1-p) \tag{6}$$

- The number T of trials up to and including the first "success":

$$f_T(n) = \text{GEOMETRIC}[p] = p(1-p)^{n-1} \tag{7}$$

$$\mathbf{E}[T] = \frac{1}{p} \tag{8}$$

$$\text{var}(T) = \frac{1-p}{p^2} \tag{9}$$

- The time Y_k until the kth success: this is just the sum of the first k interarrival times $T_1, \ldots, T_k$ which are i.i.d. geometric random variables. Y_k is distributed according to the "Pascal PMF (probability mass function) of order k":

$$f_{Y_k}(n) = \text{PASCAL}[p, k] = \binom{t-1}{k-1} p^k (1-p)^{t-k} \tag{10}$$

$$\mathbf{E}[Y_k] = \mathbf{E}\left[\sum_{i=1}^{k} T_i\right] = \frac{k}{p} \tag{11}$$

$$\text{var}(Y_k) = \text{var}\left(\sum_{i=1}^{k} T_i\right) = \frac{k(1-p)}{p^2} \tag{12}$$

2.2 Poisson approximation to the binomial distribution

Given a random variable $Z \sim \text{POISSON}[\lambda]$:

$$f_Z(k) = \text{POISSON}[\lambda] = e^{-\lambda} \frac{\lambda^k}{k!} \tag{13}$$

$$\mathbf{E}[Z] = \lambda \tag{14}$$

$$\text{var}(Z) = \lambda \tag{15}$$

[1] Bernoulli trials and Poisson trials: for a sequence of Bernoulli trials, $P(X_i = 1) = p$ and $P(X_i = 0) = 1 - p$ for some fixed p. A single Poisson trial is itself just a Bernoulli trial, but when grouped into a sequence, Poisson trials are different in that each trial may have *different* probabilities, i.e. $P(X_i = 1) = p_i$ and $P(X_i = 0) = 1 - p_i$.

Claim 2.1. If n is large and p is small, then $\text{POISSON}[np] \sim \text{BINOMIAL}[n, p]$.

Proof: Let $p = \lambda/n$. We begin with some manipulation of the binomial distribution:

$$\begin{aligned}
\text{BINOMIAL}[n, p] &= \binom{n}{k} p^k (1-p)^{n-k} && (16)\\
&= \frac{n!}{(n-k)!k!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} && (17)\\
&= \prod_{i=0}^{k-1} \frac{n-i}{n} \cdot \frac{\lambda^k}{k!} \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k} && (18)\\
& && (19)
\end{aligned}$$

Examining each component, we see that:

$$\begin{aligned}
\lim_{n\to\infty} \prod_{i=0}^{k-1} \frac{n-i}{n} &= 1 && (20)\\
\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{n-k} &= \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{n} \cdot \left(1 - \frac{\lambda}{n}\right)^{-k} && (21)\\
&= 1 \cdot e^{-\lambda} && (22)
\end{aligned}$$

(The last equality follows from the identity $\lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n = e^x$.) So we have that

$$\lim_{n\to\infty} \text{BINOMIAL}[n, p] = e^{-\lambda} \frac{\lambda^k}{k!} = \text{POISSON}[\lambda] \quad (23)$$

□

In general, the Poisson approximation is valid to within several decimal places if $n \geq 100$ and $p \leq 0.01$.

2.3 Poisson process

The Poisson process is the continuous-time analog of the Bernoulli process. Let $P(k, \tau)$ be the probability that k arrivals occur during an interval of length τ. Then a process is Poisson if it meets the following properties:

- *Time-homogeneity*: $P(k, \tau)$ is the same for any interval of length τ
- *Independence*: the number of arrivals during an interval is independent of the history of arrivals outside the interval
- *Small interval probabilities*: $P(k > 1, \tau)$ is negligible in comparison to $P(0, \tau)$ and $P(1, \tau)$ as τ decreases.

We examine the last property further. If a process is Poisson, then the PDF describing the number of arrivals in a length-τ interval is:

$$P(k, \tau) = \text{POISSON}[\lambda\tau] = e^{-\lambda\tau} \frac{(\lambda\tau)^k}{k!} \quad (24)$$

Taking the Taylor expansion for different values of k, we see that:

$$\begin{aligned}
P(0, \tau) &= e^{-\lambda\tau} \overset{\text{Taylor}}{=} 1 - \lambda\tau + O(\tau^2) && (25)\\
P(1, \tau) &= e^{-\lambda\tau}\lambda\tau \overset{\text{Taylor}}{=} \lambda\tau + O(\tau^2) && (26)\\
P(k > 1, \tau) &= O(\tau^2) && (27)
\end{aligned}$$

which verifies the final property stated above.

There are some other interesting PDFs, similar to those we examined for the Bernoulli process:

- The time T until the first arrival. Note that $T > t$ if and only if there were no arrivals in $[0, t]$, so:

$$\begin{aligned} F_T(t) &= P(P \leq t) = 1 - P(T > t) = 1 - P(0, t) = 1 - e^{-\lambda t} && (28) \\ f_T(t) &= \frac{dF_T}{dt} = \lambda e^{-\lambda t} = \text{EXPONENTIAL}[\lambda] && (29) \\ \mathbf{E}[T] &= \frac{1}{\lambda} && (30) \\ \text{var}(T) &= \frac{1}{\lambda^2} && (31) \end{aligned}$$

- The time Y_k until the kth arrival. This is again just the sum of the first k interarrival times, so Y_k follows the "Erlang PDF of order k," i.e.:

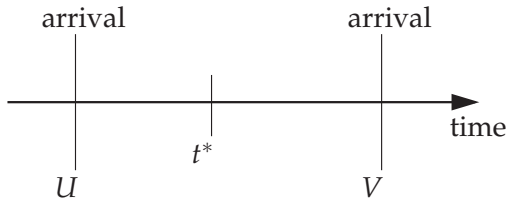
$$\begin{aligned} f_{Y_k}(y) &= \text{ERLANG}[\lambda, k] = \frac{\lambda^k y^{k-1} e^{-\lambda y}}{(k-1)!} && (32) \\ \mathbf{E}[Y_k] &= \frac{k}{\lambda} && (33) \\ \text{var}(Y_k) &= \frac{k}{\lambda^2} && (34) \end{aligned}$$

2.4 Random incidence paradox

Suppose we have a Poisson process and we fix some time t^*. Then what is the length L of the interarrival interval[2] that contains t^*? Note that t^* is not a random varaible, but L is. We'll assume that the process has been running long enough that there have been previous arrivals.

Intuitively, it seems that we'd expect L to be the length of a "typical" interarrival interval, i.e. $L \sim \text{EXPONENTIAL}[\lambda]$. However, this is not the case!

Consider the following graphical depiction:



At time U, the previous arrival occurs, and at time V, the next arrival occurs. The time we've fixed, t^*, falls between U and V. Clearly, the length of the interarrival interval in which t^* falls is then:

$$L = (t^* - U) + (V - t^*) \tag{35}$$

Because of the memorylessness property of a Poisson process, $(t^* - U)$ and $(V - t^*)$ are independent random variables, each distributed according to an exponential distribution with parameter λ. (Note that when running a Poisson process "backwards" in time, it

[2]Interarrival interval: the interval between two consecutive arrivals.

also remains Poisson.) Thus, L is the sum of two exponential random variables, i.e. $L \sim \text{ERLANG}[\lambda, 2]$ and $\mathbf{E}[L] = 2/\lambda$.

The intuition behind this apparent paradox is that an arrival at an arbitrary time is more likely to fall in a large interval rather than a small one, so the expected length of the interval in which it falls is longer than the average interarrival time.

3 Markov chains

We now consider stochastic processes where the value at some time takes on one of a finite set of *states*. Together with a model of the transitions between states and a few other properties, we have processes known as *Markov chains*.

Definition 3.1. In a *discrete-time Markov chain*, transitions require unit time. Such a chain has the following properties:

- A set of states $S = \{1, \ldots, m\}$
- A matrix P of transition probabilities, with:
  - $P_{ij} = P(X_{n+1} = j | X_n = i)$
  - $\sum_{j=1}^{m} P_{ij} = 1$ for all i
- The *Markov property* is met:

$$P(X_{n+1} = x | X_{0:n}) = P(X_{n+1} = x | X_n) \tag{36}$$

Note that given a distribution for the initial state, we can compute a distribution for the history $X_{0:n}$ of the Markov chain. For some history $x_{0:n}$:

$$P(X_0 = x_0, \ldots, X_n = x_n) = \prod_{i=0}^{n} P_{x_i x_{i+1}} \tag{37}$$

Definition 3.2. We define the *n-step transition probability* as the distribution of the state at some future time, conditioned on the current state. We specify this as:

$$\begin{aligned}
r_{ij}(n) &= P(X_n = j | X_0 = i) && (38)\\
&\overset{\text{Total prob.}}{=} \sum_{k=1}^{m} P(X_{n-1} = k | X_0 = i) P(X_n = j | X_{n-1} = k, X_0 = i) && (39)\\
&\overset{\text{Markov}}{=} \sum_{k=1}^{m} P(X_{n-1} = k | X_0 = i) P(X_n = j | X_{n-1} = k) && (40)\\
&= \sum_{k=1}^{m} r_{ik}(n-1) P_{kj} && (41)
\end{aligned}$$

Equation 41 is known as the *Chapman-Kolmogorov Equation*.

3.1 Properties of states and chains

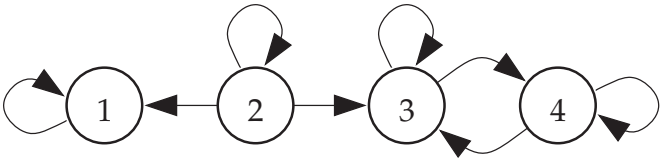
Definition 3.3. A state j is *accessible* from i if there is some n such that $r_{ij} > 0$. (It is possible to reach j from i in n steps.) We define the set $A(i)$ as:

$$A(i) \triangleq \{j \in S \mid j \text{ is accessible from } i\} \tag{42}$$

Definition 3.4. A state i is *recurrent* if for all $j \in A(i)$ it is true that $i \in A(j)$. In other words, given enough time, we will always return to i; it is only possible from i to reach states where there is some probability of returning to i.

Definition 3.5. A state i is *transient* if it is not recurrent. Given enough time, we will eventually enter a state from which it is impossible to return to i. Thus, i will be visited a finite number of times.

The following picture depicts recurrent and transient states:

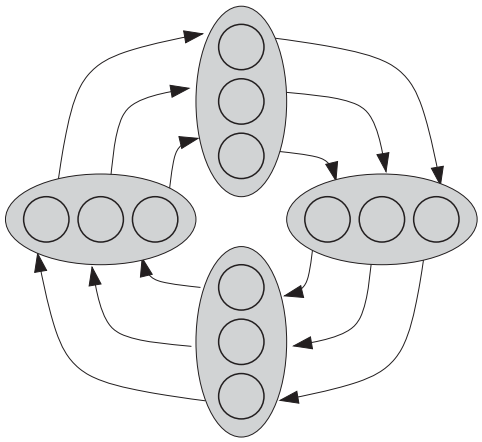


States 1, 3, and 4 are recurrent, while state 2 is transient.

Definition 3.6. If state i is recurrent, $A(i)$ forms a *recurrent class*. Then, all states in $A(i)$ are accessible from each other and no states outside $A(i)$ are accessible from a state in $A(i)$.

A Markov chain can be decomposed into one or more recurrent classes, plus some transient states. Once the process enters a recurrent class, the process remains in that class forever and all states in the class are visited an infinite number of times. If X_0 is transient, then for some k, $X_{0:k}$ are all transient and $X_{k+1:\infty}$ are all in the same recurrent class.

Definition 3.7. Let R be some recurrent class. R is *periodic* if its states can be grouped into $d > 1$ disjoint subsets $S_1, \ldots, S_d$ such that all transitions from S_k lead to S_{k+1} and all transitions from S_d lead to S_1. In other words, the states in R can be partitioned like the following:



A recurrent class that is not periodic is termed *aperiodic*.

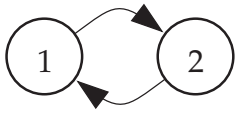
Definition 3.8. An *ergodic state* is one that is both aperiodic and recurrent.

Definition 3.9. An *ergodic Markov chain* is one in which all states are ergodic.

3.2 Steady-state behavior

It is frequently useful to analyze the long-term state occupancy behavior of a Markov chain. In other words, what is the behavior of $r_{ij}(n)$ as n becomes large?

First, note that if there are multiple recurrent classes, $\lim_{n\to\infty} r_{ij}(n)$ depends on the initial state. Furthermore, even if there is a single recurrent class, $r_{ij}(n)$ may not converge if the class is periodic, as seen in the following simple example:



Here, regardless of the length of the process, we have:

$$r_{ii}(n) = \begin{cases} 1 & \text{if } n \text{ is even} \\ 0 & \text{if } n \text{ is odd} \end{cases} \tag{43}$$

Definition 3.10. For these reasons, we restrict our study of convergence to Markov chains with a single aperiodic recurrent class. For such a Markov chain, we have:

$$\lim_{n\to\infty} r_{ij}(n) = \pi_j \approx P\left(X_n = j\right) \tag{44}$$

Here, π_j is the *steady-state probability of j*.

Definition 3.11. Combining the above with the Chapman-Kolmogorov Equation:

$$\lim_{n\to\infty} \sum_{k=1}^{m} r_{ik}(n-1)P_{kj} = \pi_j \tag{45}$$

$$\sum_{k=1}^{m} \pi_k P_{kj} = \pi_j \tag{46}$$

The equations 46 are known as the *balance equations*.

We can consider the balance equations together with the normalization equation $\sum_{k=1}^{m} \pi_k = 1$ and form a system of equations, which can be solved to obtain the π_j's. A property of the solution is that $\pi_j = 0$ if j is transient, and $\pi_j > 0$ otherwise.

The steady-state probabilities can be thought of as *expected state frequencies*. For a Markov chain with a single aperiodic recurrent class,

$$\pi_j = \lim_{n\to\infty} \frac{v_{ij}(n)}{n} \tag{47}$$

where $v_{ij}(n)$ is the expected value of the number of visits to j within the first n transitions, starting from i.

We can also examine the *frequency of transitions*. Let $q_{jk}(n)$ be the expected number of times the transition from j to k is taken in the first n steps. Then:

$$\lim_{n\to\infty} \frac{q_{jk}(n)}{n} = \pi_j P_{jk} \tag{48}$$

3.3 Absorption

It is also useful to analyze the short-term behavior of Markov chains. An important question is: if we start in a transient state, how long will it be until we enter the first recurrent state? Since what happens afterward does not matter, we focus on the case where every recurrent state is *absorbing*, i.e. $P_{kk} = 1$.

Definition 3.12. Fix some absorbing state s. The probability that s is eventually reached, starting from state i, is termed the *absorption probability*:

$$a_i = P\left(X_n \text{ eventually becomes } s | X_0 = i\right) \tag{49}$$

Clearly, we have $a_s = 1$. Furthermore, for all absorbing $i \neq s$, $a_i = 0$. The remaining case is that in which i is transient. Let A be the event that s is eventually reached. For a transient state i, we have:

$$\begin{aligned} a_i &= P\left(A | X_0 = i\right) && (50)\\ &= \sum_{j=1}^{m} P\left(A | X_0 = i, X_1 = j\right) P\left(X_1 = j | X_0 = j\right) && (51)\\ &\overset{\text{Markov}}{=} \sum_{j=1}^{m} P\left(A | X_1 = j\right) P_{ij} && (52)\\ &= \sum_{j=1}^{m} a_j P_{ij} && (53) \end{aligned}$$

The equations for a_i form a system which we can solve to obtain the absorption probabilities.

There are several other related properties that we can examine.

Definition 3.13. *Expected time to absorption*: the time until we enter some recurrent state. Let

$$\begin{aligned} \mu_i &= \mathbf{E}\left[\text{num. transitions until a recurrent state is entered} | X_0 = i\right] && (54)\\ &= \mathbf{E}\left[\min\{n \geq 0 | X_n \text{ is recurrent}\} | X_0 = i\right] && (55) \end{aligned}$$

Then:

$$\mu_i = \begin{cases} 0 & \text{if } i \text{ is recurrent} \\ 1 + \sum_{j=1}^{m} P_{ij}\mu_j & \text{if } i \text{ is transient} \end{cases} \tag{56}$$

These equations can be solved to find the unique expected time to absorption.

Definition 3.14. *Mean first passage time* t_i: the time until we reach recurrent state s, starting from i.

$$t_i = \begin{cases} 0 & \text{if } i = s \\ 1 + \sum_{j=1}^{m} P_{ij} t_j & \text{otherwise} \end{cases} \tag{57}$$

These equations can again be solved to find the unique mean first passage time.

Definition 3.15. *Mean recurrence time* t_s^*: the number of transitions up to the first return to s, starting from s:

$$t_s^* = 1 + \sum_{j=1}^{m} P_{sj} t_j \tag{58}$$

3.4 Continuous-time Markov chains

Definition 3.16. In a *continuous-time Markov chain*, the time between transitions is a continuous random variable. We define the following random variables:

$$X_n \triangleq \text{state after the } n\text{th transition} \tag{59}$$
$$Y_n \triangleq \text{time of the } n\text{th transition} \tag{60}$$
$$T_n \triangleq \text{time between the } (n-1)\text{st and } n\text{th transitions} \tag{61}$$

Definition 3.17. For some state i, the time T until the next transition is distributed according to $\text{EXPONENTIAL}[\nu_i]$ where ν_i is the *transition rate*, i.e. the average number of transitions out of i per unit time spent in i. We assume that T is independent of the past history of the process and the next state.

Definition 3.18. A continuous-time Markov chain defines the transition probability matrix as before. Combining this with the transition rate, we define $q_{ij} = \nu_i P_{ij}$ to be the *transition rate* from i to j, i.e. the average number of transitions from i to j per unit time spent in i.

As essentially the same steady-state convergence results apply for continuous-time Markov chains as for discrete-time Markov chains, we omit further discussion.