Stochastic processes

AMINA BENACEUR

February 4, 2025

1 Preliminaries

Probability theory studies random phenomena. When observing a random phenomenon, the set of possible outcomes ω lies in a sample space Ω. For technical reasons, the collection of events $\mathcal{F}$ assigned a probability is larger than the collection of all subsets of Ω, i.e.

- The impossible event $\emptyset$ and the certain event Ω are in $\mathcal{F}$;

- If A is in $\mathcal{F}$, the so is $\bar{A}$;

- The union of events in $\mathcal{F}$ is also in $\mathcal{F}$.

1.1 Random variables

A random variable (rv) is roughly speaking a numerical quantity that takes random values. Consider a set of students $s_1, s_2, \ldots, s_n$ in this class. We record their sleep time that we denote X. X is a random function that takes different random values for each student. We can do the same for their wake up time W. Wa plug in the student to the function, and get the numerical value x or Y. X and Y are abstract object/quantity whose value is determined once we know the outcome of the experiment.

Definition 1. *A random variable is a function $X : \Omega \to \mathbb{R}$ such that for all $a \in \mathbb{R}$, the event $\{X \leq a\} = \{\omega; X(\omega) \leq a\}$ can be assigned a probability, that is, $\{X \leq a\} \in \mathcal{F}$. A function X for which E is denumerable is called a discrete random variable if for all $x \in E$, $\{X = x\} \in \mathcal{F}$.*

In plain words, a random variable is a variable whose possible values are numerical outcomes of a random phenomenon. There are two types of random variables, discrete and continuous.

- A discrete random variable is one which may take on only a countable number of distinct values. Discrete random variables are usually counts. If a random variable can take only a finite number of distinct values, then it must be discrete. Examples of discrete random variables include

  1. The number of children in a nuclear family;
  2. The number of students in an on-campus food truck;
  3. A prayer attendance at the UM6P mosque;

4. The number of professors in a department.

The probability distribution (or probability mass function) of a discrete random variable is a list of probabilities associated with each of its possible values.

- A continuous random variable is one which takes an infinite number of possible values. Continuous random variables are usually measurements.

  Examples of continuous random variables include

  1. Height;
  2. Weight;
  3. Water flow rate in a sink.

  Examples include height, weight, the amount of sugar in an orange, the time required to run a mile.

Remark 1. *1. Several rvs can be defined on the same sample space.*

2. A function of one or more rvs is also an rv.

If X is a discrete r.v., then thefinite or countably in?nite set of values x such that $P(X = x) > 0$ is called the support of X. Discrete random variables have a probability mass function (PMF) which gives the probability of an event $X = a$ as $P(X = a)$. Note that this is positive if x is in the support of X, and 0 otherwise. The support of a discrete r.v. is a set of integers. In contrast, a continuous r.v. can take on any real value in an interval (possibly even the entire real line).

Example 1. *1. Consider X to be the outcome of a fair coin toss. What is the support of X? What is $P(X = H)$ and $P(X = T)$?*

2. Consider Y to be the outcome of a fair die roll. What is the support of Y. What is $P(Y = 6)$ and $P(Y = T)$?

Theorem 1 (Valid PMFs). *. Let X be a discrete r.v. with support $x_1, x_2, \ldots$ (assume these values are distinct and, for notational simplicity, that the support is countably infinite; the analogous results hold if the support is finite). The PMF p_X of X must satisfy the following criteria:*

1. Nonnegative: $p_X(x) > 0$ if x is in the support of X and $p_X(x) = 0$ otherwise;

2. Sums to 1: $\sum_{j=1}^{\infty} p_X(x_j) = 1$.

For continuous variables $P(X = a) = 0$.

All random variables (discrete and continuous) have a cumulative distribution function (CDF). It is a function giving the probability that the random variable X is less than or equal to x, for every value x, i.e., $F_X : x \mapsto P(X \leq x)$. For a discrete rv, the cumulative distribution function is found by summing up the PMF, i.e., $\sum_{x_k \leq x} p(x_k)$. Fo a continuous rv, it is given by intergrating the PDF, i.e. $\int_{-\infty}^{x} f$.

Exercise 1.   *1. Benford?s law states that in a very large variety of real-life data sets, the first digit approximately follows a particular distribution with about a 30% chance of a 1, an 18% chance of a 2, and in general*

$$P(D = j) = log_{10}\left(\frac{j+1}{j}\right), \quad \forall j \in \{1, 2, 3, \ldots, 9\}, \tag{1}$$

where D is the first digit of a randomly chosen element. Check that this is a valid PMF.

1.2   Bernoulli, binomial, exponential, Indicator rv, Poisson

Some distributions have so many applications that they have their own names.

We start with a very simple but useful case: an r.v. that can take on only two possible values, 0 and 1.

Definition 2 (Bernoulli distribution). *. An r.v. X is said to have the Bernoulli distribution with parameter p if $P(X = 1) = p$ and $P(X = 0) = 1 - p$, where $0 < p < 1$. We write this as $X \sim \text{Bern}(p)$. The symbol $\sim$ is read "is distributed as".*

We often imagine Bernoulli rvs using coin tosses, with $X = 1$ if heads and $X = 0$ if tails. An experiment that can result in either a "success" or a "failure" is called a Bernoulli trial. A Bernoulli rv can be thought of as the indicator of success in a Bernoulli trial: it equals 1 if the trial succeeds and 0 if the trial fails. The parameter p is generally called the success parameter or the probability of success.

Imagine the case where we have a set of n independent Bernoulli trials with the same parameter p, a coin toss as an example. Let X be the number of successes. The distribution of X is called the binomial distribution with parameters n and p. We thus write $X \sim \text{Bin}(n, p)$. Note that both the Bernoulli and the binomial distributions can be connected to the stories/experiments that initiated them. It is good to keep them in mind to develop intuition that often eases problem solving.

Proposition 1. *The PMF of a binomial distribution $\text{Bin}(n, p)$ is*

$$P(X = k) = \binom{k}{n} p^k (1 - p)^{n-k} \tag{2}$$

for $0 \le k \le n$, and 0 for $k > n$.

Definition 3 (Binomial distribution). *. An r.v. X is said to have the Bernoulli distribution with parameter p if $P(X = 1) = p$ and $P(X = 0) = 1 - p$, where $0 < p < 1$. We write this as $X \sim \text{Bern}(p)$. The symbol $\sim$ is read "is distributed as".*

Exercise 2. *Let $X \sim \text{Bin(n, p)}$, and $q = 1 - p$.*

1. *What is p in terms of probability? (probability of failure)*

2. *We introduce the rv $Y = n - X$. What is the PMF of Y $(Bin(n, q))$?*

3. *Independent Bernoulli trials are performed, with probability 1/2 of success, until there has been at least one success. Find the PMF of the number of trials performed.*

4. *Independent Bernoulli trials are performed, with probability 1/2 of success, until there has been at least one success and at least one failure. Find the PMF of the number of trials performed.*

1.3   Probability inequalities

The Cauchy-Schwarz inequality is one of the most famous inequalities in all of mathematics. Its probabilistic form is as follows

Theorem 2 (Cauchy-Schwarz). *For any r.v.s X and Y with finite variances,*

$$|E[XY]| \leq \sqrt{E[X^2]E[Y^2]}. \tag{3}$$

Proof.

$$\forall t \in \mathbb{R}, \quad 0 \leq E[Y - tX]^2 = E[Y^2] - 2tE[XY] + t^2 E[X^2] \tag{4}$$

Differentiate with respect to t and see where the minimum is. $\qquad\square$

Example 2. *Let X, Y be two rvs such that $E[X] = E[Y] = 0$. Prove that $-1 \leq cov(X, Y) \leq 1$.*

Theorem 3 (Markov inequality). *For any rv X and a constant $a > 0$*

$$P(X > a) \leq \frac{E[X]}{a} \tag{5}$$

For an intuitive interpretation, let X be the age of a randomly selected individual from a population. Taking $a = 2E[X]$, Markov?s inequality says that $P(X \geq 2E[X]) \leq 0.5$, i.e., it is impossible for more than half the population to be at least twice the average age. This is clearly true, since if over half the population were at least twice the average age, those people would already drive the average above what it is.

Chebychev is a prominent Russian mathematician fom the 19th century.

Theorem 4 (Chebychev inequality). *Let X be an rv with mean μ and variance σ^2.*

$$\forall a > 0, \quad P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2} \tag{6}$$

Proof. Markov's property yields

$$P\left((X - \mu)^2 > a^2\right) \leq \frac{E[(X - \mu)^2]}{a^2}.$$

$P\left((X - \mu)^2 > a^2\right) = P\left(|X - \mu| > a\right)$ and $E[(X - \mu)^2] = \sigma^2$, whereof the result. $\qquad\square$

For $a = c\sigma$, $c > 0$, we get

$$P(|X - \mu| \geq c\sigma) \leq c^2.$$

To get an intuition of the inequality, it gives us an upper bound on the probability of an r.v. being more than c standard deviations away from its mean, e.g., there can't be more than a 25% chance of being 2 or more standard deviations from the mean.

1.4   Limit theorems

Let $X_1, \ldots, X_n$ be i.i.d random variables with mean μ and variance σ^2. Let

$$\bar{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

The variance of $\bar{X}_n$ is σ^2/n.

1.4.1 Weak Law of large numbers (WLLN)

Theorem 5. *For $n \geq 1$, and i.i.d random variables $X_1, \ldots, X_n$ with finite variance, it holds that*

$$\forall \epsilon \geq 0, \qquad \lim_{n \to \infty} Pr\left\{ \left| \left(\bar{X}_n - \mu \right) \right| \geq \epsilon \right\} = 0. \tag{7}$$

Proof. Using Chebychev inequality, we get

$$Pr\left(\left| \bar{X}_n - \mu \right| \geq \epsilon \right) \leq \frac{\sigma^2}{n\epsilon^2} \underset{n \to \infty}{\to} 0.$$

$\square$

Every time we use the proportion of times that something happened as an approximation to its probability, we are implicitly appealing to the LLN. Every time we use the average value in the replications of some quantity to approximate its theoretical average, we are implicitly appealing to the LLN.

Remark 2 (LLN does not contradict the memoryless property). *. In the above exam- ple, the law of large numbers states that the proportion of Heads converges to 1/2, but this does not imply that after a long string of Heads, the coin is due? for a Tails to balance things out. Rather, the convergence takes place through swamping: past tosses are swamped by the infinitely many tosses that are yet to come*

1.4.2 Central limit theorem (CLT)

The law of large numbers says that as $n \to \infty$, $\bar{X}_n$ converges to the constant μ (with probability 1). But what is its distribution along the way to becoming a constant? This is addressed by the central limit theorem (CLT). The CLT states that for large n, the distribution of $\bar{X}_n$ after standardization approaches a standard normal distribution $\mathcal{N}(0, 1)$. By standardization, we mean that we subtract μ, the expected value of $\bar{X}_n$, and divide by $\sigma/\sqrt{n}$, the standard deviation of $\bar{X}_n$.

Theorem 6 (Central limit theorem). *Let $X_1, \ldots, X_n$ be i.i.d random variables with finite mean μ and finite variance σ^2. It holds that*

$$\lim_{n \to \infty} Pr\left\{ \frac{\bar{X}_n - \mu}{\sigma \sqrt{n}} \leq z \right\} = \mathcal{N}(0, 1)(z) \tag{8}$$

which is referred to as convergence in distribution

Let?s take a moment to admire the generality of this result. The distribution of the individual X_i can be anything in the world, as long as the mean and variance are finite. We could have a discrete distribution like the Binomial, a bounded distribution like the Beta, or a distribution with multiple peaks and valleys. No matter what, the act of averaging will cause Normality to emerge.

This does not mean that the distribution of the Xj is irrelevant, however. If the X_i have a highly skewed or multimodal distribution, we may need n to be very large before the Normal approximation becomes accurate; at the other extreme, if the X_i are already i.i.d. Normals, the distribution of $\bar{X}_n$ is exactly $\mathcal{N}(\mu, \sigma^2/n)$ for all n.

1.5 Stochastic processes

Definition 4. *We give two definitions*

- *A stochastic process is a sequence of random variables with values in a set E indexed by time. The space E is called a state space. E is a countable space.*

- *A stochastic process is a probability distribution over a set of paths*

Examples:

1. Discrete-time stochastic process $(X_i)_{1 \leq i \leq n}$. If $X_n = i$, then the process is said to be in state i at time n.

2. Continuous time stochastic-process $(X_t)_{t \geq 0}$.

Examples of stochastic processes

1.5.1 Random walk

Consider a set of i.i.d variables $Y_1, \ldots, Y_n$ where for each variable Y_k, $k \in \{1, \ldots, n\}$, $Y_k \in \{-1, 1\}$ with equal probability. $X_1, X_2, X_3, \ldots$ is a stochastic process, and we introduce a random walk as

$$X_N = Y_1 + Y_2 + \ldots + Y_N. \tag{9}$$

By the central limit theorem, we know that for a large enough integer n, the distribution of $\frac{1}{\sqrt{n}} X_n$ converges to the normal distribution with mean 0 and variance 1. This already tells us some information about the random walk. We state some further properties of the random walk

Proposition 2. *The following statements hold true*

1. $\forall k \geq 0, \ E[X_k] = 0.$

2. *The random walk has independent increments, i.e., $\forall \ 0 = k_0 \leq k_1 \leq \ldots \leq k_r$, the r.v.s $X_{k_{i+1}} - X_{k_i}$ for $0 \leq i \leq r-1$ are mutually independent.*

3. *The random walk is stationary, i.e., $\forall, k, h \geq 0$, the distribution of $X_{k+h} - X_k$ is the same as the distribution of X_h.*

Exercise: A particle moves n steps on a number line. The particle starts at 0, and at each step it moves 1 unit to the right or to the left, with equal probabilities. Assume all steps are independent. Let Z be the particle?s position after n steps. Find the PMF of Z. If k is an integer between ?n and n (inclusive) such that n+k is an even number, we get

$$P(Z = k) = \binom{n}{\frac{n+k}{2}} \left(\frac{1}{2}\right)^n$$

otherwise $P(Z = k) = 0$

For two positive integers A and B, what is the probability that the random walk reaches A before it reaches $-B$?

1.6 Application to Fair Gambling Games

We took the high road for a while, but let?s now discuss random walks in a more conventional setting?gambling.

A gambler goes to Las Vegas with n in her pocket. Her plan is to make only 1 bets and somehow she has found a casino that will offer her truly even odds; namely, she will win or lose 1 on each bet with probability $\frac{1}{2}$. She?ll play until she is broke or she has won m. In the latter case, she will go home with

$$w = n + m$$

dollars. What?s the probability that she goes home a winner?

This is identical to the flea problem that we just analyzed. Going broke is analogous to falling off the Cliff of Doom. Going home a winner is analogous to falling into the Pit of Disaster, just a lot more fun.

Our analysis of Stencil?s life tells us everything we want to know about the gambler?s prospects:

- The gambler goes home broke with probability

$$\frac{n}{n+m};$$

- The gambler goes home a winner with probability

$$\frac{m}{n+m};$$

- The gambler goes home with probability

$$\frac{n}{n+m} + \frac{m}{n+m} = 1;$$

- And the number of bets before the gambler goes home is expected to be

$$n \cdot (w - n) = nm.$$

If the gambler gets greedy and keeps playing until she goes broke, then

- The gambler eventually goes broke with probability 1, and

- The number of bets before the gambler goes broke is expected to be infinite.

The bottom line here is clear: when gambling, quit while you are ahead?if you play until you go broke, you will certainly go broke. And that?s the good news!

Matters get much worse for the more typical scenario where the odds are against you. Let?s see why.

1.6.1 Gambler?s Ruin

So far, we have considered unbiased random walks, where the probability of moving up or down (or left or right) is $\frac{1}{2}$. Unfortunately, things are never quite this simple (or fair) in real casinos.

For example, suppose the gambler goes to Las Vegas and makes 1 bets on red or black in roulette. In this case, she will win 1 with probability

$$\frac{18}{38} \approx 0.473$$

and she will lose 1 with probability

$$\frac{20}{38} \approx 0.527.$$

That?s because the casinos add those bothersome green 0 and 00 to give the house a slight advantage.

At first glance (or after a few drinks), $\frac{18}{38}$ seems awfully close to $\frac{1}{2}$ and so our intuition tells us that the game is ?almost fair.? So we might expect the analysis we just did for the fair game to be ?almost right? for the real game. For example, if the gambler starts with 100 and quits when she gets ahead by 100 in the fair game, then she goes home a winner with probability

$$\frac{100}{200} = 0.5.$$

And, if she wants to improve her chances of going home a winner, she could bring more money. If she brings 1000 and quits when she gets ahead by 100 in the fair game, then she goes home a winner with probability

$$\frac{1000}{1100} \approx 0.91.$$

So, given that the real game is ?almost fair,? we might expect the probabilities of going home a winner in these two scenarios to be ?almost 50% and 91%,? respectively.

Unfortunately for the gambler, all this ?almost? reasoning will almost surely lead to disaster. Here are the grim facts for the real game where the gambler wins 1 with probability $\frac{18}{38}$.

For $n = 100$, probability she reaches $n + 100$ before 0 : 1 in 37649.619496...

For $n = 1000$, probability she reaches $n + 100$ before 0 : 1 in 37648.619496...

Except on the very low end, the amount of money she brings makes almost no difference! She is almost certain to go broke before winning 100. Let?s see why.

1.6.2 Finding a Recurrence

We can approach the gambling problem the same way we studied the life of Stencil. Suppose that the gambler starts with n dollars. She wins each bet with probability p and plays until she either goes bankrupt or has $w = n + m$ dollars in her pocket. (To be clear, w is the total amount of money she wants to end up with, not the amount by which she wants to increase her wealth, which is m.) Our objective is to compute R_n, the probability that she goes home a winner.

As usual, we begin by identifying some boundary conditions. If she starts with no money, then she?s bankrupt immediately so $R_0 = 0$. On the other hand, if she starts with w dollars, then she?s an instant winner, so $R_w = 1$.

Now we divide the analysis of the general situation into two cases based on the outcome of her first bet:

- She wins her first bet with probability p. She then has $n+1$ dollars and probability R_{n+1} of reaching her goal of w dollars.

- She loses her first bet with probability $1-p$. This leaves her with $n-1$ dollars and probability R_{n-1} of reaching her goal.

Plugging these facts into the Total Probability Theorem gives the equation:

$$R_n = pR_{n+1} + (1-p)R_{n-1}.$$

1.6.3 Solving the Recurrence

Rearranging the terms in the equation gives us a recurrence for R_n, the probability that the gambler reaches her goal of w dollars if she starts with n:

$$R_0 = 0,$$

$$R_w = 1,$$

$$pR_{n+1} - R_n + (1-p)R_{n-1} = 0 \quad (0 < n < w).$$

The characteristic equation is:

$$px^2 - x + (1-p) = 0.$$

Using the quadratic formula, we get the roots:

$$x = 1 \pm \frac{\sqrt{1 - 4p(1-p)}}{2p}.$$

If the gambler is equally likely to win or lose each bet, then $p = \frac{1}{2}$, and the characteristic equation has a double root at $x = 1$. This is the situation we considered in the flea problem. The general solution in this case is:

$$R_n = a + bn.$$

Suppose that the gambler is not equally likely to win or lose each bet; that is, $p \neq \frac{1}{2}$. Then the two roots of the characteristic equation are different, leading to the general solution:

$$R_n = a\left(\frac{1-p}{p}\right)^n + b \times 1^n.$$

In mathematical terms, this is where the fair game and the "almost fair" game take off in completely different directions: in one case we get a linear solution and in the other we get an exponential solution! This is going to be bad news for anyone playing the "almost fair" game.

Anyway, substituting the boundary conditions into the general form of the solution gives a system of linear equations:

$$0 = a + b$$

$$1 = a\left(\frac{1-p}{p}\right)^w + b$$

Solving this system gives:

$$a = \frac{1}{\left(\frac{1-p}{p}\right)^w - 1}, \quad b = -\frac{1}{\left(\frac{1-p}{p}\right)^w - 1}$$

Substituting these values back into the general solution gives:

$$R_n = \left(\frac{1-p}{p}\right)^{w-1} \left(\frac{1-p}{p}\right)^{n-1}$$

(Suddenly, Stencil's life doesn't seem so bad, huh?)

Bad news! We have an answer! But it's not good news. If the gambler starts with n dollars and wins each bet with probability p, then the probability she reaches w dollars before going broke is:

$$\frac{\left(\frac{1-p}{p}\right)^n - 1}{\left(\frac{1-p}{p}\right)^w - 1}$$

Let's try to make sense of this expression. If the game is biased against her, as with roulette, then $1 - p$ (the probability she loses) is greater than p (the probability she wins). If n, her starting wealth, is also reasonably large, then both exponentiated fractions are big numbers and the -1's don't make much difference. Thus, her probability of reaching w dollars is very close to:

$$\left(1 - \frac{1}{p}\right)^{n-w} = \left(1 - \frac{1}{p}\right)^m$$

In particular, if she is hoping to come out $m = 100$ dollars ahead in roulette, then $p = \frac{18}{38}$ and her probability of success is:

$$\left(\frac{10}{9}\right)^{-100} = \frac{1}{37648.619496}$$

This explains the strange number we arrived at earlier! In fact, this number does not change no matter how large n gets, so even if the gambler starts with a trillion dollars, she is still not likely to ever get ahead by even 100.

But Why?

Why does the gambler's starting wealth have so little impact on her probability of coming out ahead? Intuitively, there are two forces at work. First, the gambler's wealth has random upward and downward swings due to runs of good and bad luck. Second, her wealth has a steady, downward drift because she has a small expected loss on every bet. The situation is illustrated in Figure 20.2.

For example, in roulette, the gambler wins a dollar with probability $\frac{18}{38}$ and loses a dollar with probability $\frac{20}{38}$. Therefore, her expected return on each bet is:

$$1 \times \frac{18}{38} + (-1) \times \frac{20}{38} = \frac{2}{38} = \frac{1}{19}$$

Thus, her expected wealth drifts downward by a little over 5 cents per bet.

One might think that if the gambler starts with a billion dollars, then she will play for a long time, so at some point she should have a lucky, upward swing that puts her 100 ahead.

The problem is that her capital is steadily drifting downward. And after her capital drifts down a few hundred dollars, she needs a huge upward swing to save herself. And such a huge swing is extremely improbable. So if she does not have a lucky, upward swing early on, she's doomed forever. As a rule of thumb, drift dominates swings over the long term.

2 Discrete Markov chains

2.1 The Markov Property

Let X_n be a stochastic process with values in a countable state space E. (X_n) is said to be a Markov chain (MC) if and only if

$$P(X_{k+1} = j \mid X_k = i, X_{k-1} = i_{k-1}, \ldots, X_0 = i_0) = P(X_{k+1} = j \mid X_k = i), \tag{10}$$

For convenience, (11) is often abbreviated as

$$P(X_{k+1} \mid X_k, X_{k-1}, \ldots, X_0) = P(X_{k+1} \mid X_k), \tag{11}$$

which means that the equality holds independently of sample values. The property (11) is called the *Markov property*. Additionally, $(X_n)_{n \geq 0}$ is called a homogenous Markov chain (HMC) if the right-hand side of (11) is independent of n. We introduce the probabilities p_{ij} such that

$$p_{ij} = P(X_{n+1} = j \mid X_n = i) \tag{12}$$

The matrix $\mathbf{P} = (p_{ij})_{i,j \in E}$ is the *transition matrix*. The elements of $\mathbf{P}$ are probabilities. Therefore, they satisfy

$$\forall i, j \in E, \ p_{ij} \geq 0, \quad \text{and} \sum_{j \in E} p_{ij} = 1. \tag{13}$$

Definition 5. *A Markov chain is a discrete-time stochastic process $(X_k)_{k \geq 0}$ that satisfies the Markov property.*

In other words, a Markov chain is a stochastic process for which the sample values for each rv in $(X_k)_{k \geq 0}$ lie in a countable set E and depend on the past only through the most recent rv X_{k-1}

An HMC and its transition matrix $\mathbf{P}$ are usually represented by a directed graph called a *transition graph G*. The nodes of G are the states of the process. G has an oriented edge from node i to node j if and only if $p_{ij} > 0$. The transition probabilities p_{ij} are inscribed on the corresponding oriented edge. The difference between zero and non-zero transition probabilities stands out clearly on a graph as edges are not represented until transition probabilities are non-zero.

2.2 Classification of states

Definition 6.　*1. An n-step walk is an ordered set of nodes $(i_0, \ldots, i_n), n \geq 1$, for which there is a directed edge from i_{k-1} to i_k for each $k \in \{\{1, \ldots, n\}\}$.*

    2. A path is a walk in which no nodes are repeated A cycle is a walk in which the first and last nodes coincide.

    3. A state j is accessible from i f there is a walk from i to j

    4. Two distinct states i and j communicate (denoted $i \leftrightarrow j$) if i is accessible from j and vice versa.

5. *A class is a non-empty set of states such that:* $\forall i \in \mathcal{C}$, $\forall j \neq i \in \mathcal{C}$, *if* $i \leftrightarrow j$, *then* $j \in \mathcal{C}$, *and if* $i \nleftrightarrow j$, *then* $j \notin \mathcal{C}$.

6. *For finite-state Markov chains, a recurrent state is a state* i *that is accessible from* i *(i is recurrent if* $i \rightarrow j$ *implies that* $j \leftarrow i$

7. *A transient state is a state that is not recurrent.*

8. *The period* $d(i)$ *of a state* i *the greatest common divisor (gcp) of those values of* n *for which* P_{ii}^n. *If the period is 1, the state is aperiodic, and if the period is 2 or more, then the state is periodic.*

Question 1. *What is the maximum number os steps in a path? What mathematical property? What is the probability to return to a recurrent state on the long run?*

Theorem 7. *For finite-state MC, a class either consits exclusively of recurrent states, or consists exclusively of recurrent states*

Theorem 8. *For a MC, all states in the same class have the same period.*

Definition 7. *For a finite-state MC, an ergodic class of states is a class that is both recurrent and aperiodic. A MC consisting entirely of one ergodic class is called an ergodic chain.*

2.3 Distribution of an HMC

The rv X_0 is called the initial state. From Bayes' rule

$$P(X_0 = i_0, \ldots, X_k = i_k) = P(X_0 = i_0)P(X_1 = i_1 \mid X_0 = i_0)\ldots P(X_k = i_k \mid X_{k-1} = i_{k-1}) \quad (14)$$

Denoting the probability distribution of the initial state as $v(i_0) = P(X_0 = i_0)$, we obtain

$$P(X_0 = i_0, \ldots, X_k = i_k) = v(i_0)p_{i_0 i_1}p_{i_1 i_2}\ldots p_{i_{k-1} i_k} \quad (15)$$

which constritutes the distribution of the HMC.

Theorem 9. *The distribution of a discrete-time HMC is determined by its initial distribution and its transition matrix.*

2.4 Properties of Markov chains

2.5 Communicating classes

2.6 Stationarity and the steady state

Definition 8. *A probability distribution satisfying*

$$\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi} \quad (16)$$

is called a stationary distribution of the transition matrix $\mathbf{P}$. *It can also be called a stationary distribution of the corresponding HMC.*

Proposition 3. *Let π be a stationary distribution for $(X_n)_{n\geq 0}$. $\forall n \geq 0, \forall x \in E$, $Pr(X_n = x) = \pi(x)$, i.e., if thprobability distribution of the initial state X_0 is π, then for any $n \geq 0$, the probability distribution of X_n is also π.*

Proposition 4. *If a Markov chain is defined over a finite state space E, then it admits a stationary distribution*

Proof. Perron-Frobenius theorem $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Proposition 5. *If a Markov chain chain is irreducible then all states are recurrent.*

Proposition 6. *If a Markov chain chain is irreducible with finitely many states and aperiodic, then*

1. *It has a stationary distribution.*

2. *The stationary distribution is unique.*

3. *$s_i = \frac{1}{r_i}$, where r_i is the expected time to return to state i.*

Exercise 3. *Find the stationary distribution for some Markov chains*

Definition 9. *A Markov chain with transition matrix $\mathbf{P}$ on E is irreducible if, for all $x, y \in E$, there exist $n \geq 1$ and $x = x0, x_1, ..., x_n = y$ such that $P(x_0, x_1) \ldots P(x_{n?1}, x_n) > 0$.*

2.7 Periodicity

If self-transition, then the MCis not periodic
   Eyballing is not always sufficient

2.8 Time reversal

2.9 Chapman-Kolmogorov equation

Theorem 10.

$$P_{ij}(n + m) = \sum_{k=1}^{n} P_{ik}(n)P_{kj}(m)$$

Proof.

$$
\begin{aligned}
P_{ij}(n + m) &= P(X_{n+m} = j | X_0 = i) \\
&= \sum_{k=1}^{n} P(X_{n+m} = j \cap X_n = k | X_0 = i) \\
&= \sum_{k=1}^{n} P(X_{n+m} = j | X_0 = i \cap X_n = k)P(X_n = k | X_0 = i) \\
&= \sum_{k=1}^{n} P(X_{n+m} = j | X_n = k)P(X_n = k | X_0 = i) \\
&= \sum_{k=1}^{n} P(X_m = j | X_0 = k)P(X_n = k | X_0 = i) \\
&= \sum_{k=1}^{n} P_{kj}(m)P_{ik}(n),
\end{aligned}
\tag{17}
$$

14

wherof the result. □

Definition 10. *As stochastic process is reversible if and only if*

$$\pi(x)\text{Pr}(x, y) = \pi(y)\text{Pr}(y, x) \tag{18}$$

A Markov chain X is reversible with respect to a probability measure π if its transition matrix is reversible with respect to π.

Proposition 7. *If a stochastic matrix $\mathbf{P}$ is reversible with respect to a probability measure π, then π is a stationary probability measure for $\mathbf{P}$.*

Proposition 7 is derived by summing (18) over $x \in E$.
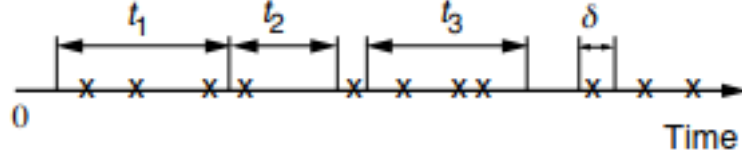
3 Poisson processes

Consider a process that consists in people randomly arriving with $P(k, \tau)$ the probability of k arrivals during an interval of length τ. Let's additionally assume that the number of arrivals in disjoint time intervals are iid. We introduce λ as the arrival rate, which corresponds to the number of arrivals in a unit of time. For a time interval of width δ small enough (i.e. at the limit $\delta \to 0$), the following approximation holds:

$$P(k, \delta) \approx \begin{cases} 1 - \lambda\delta, & \text{if } k < 1, \\ \lambda\delta, & \text{if } k = 1. \\ 0, & \text{if } k > 1 \end{cases} \tag{19}$$

Let's consider a Bernoulli process over a time interval of width τ. Let's subdivide the interval into n subintervals of length τ/n each. The probability of an arrival during each of these intervals is $p = \lambda\tau/n$. The binomial PMF giving the probability of k arrivals over n intervals for this Bernoulli process is then given by

$$P(k, n) = \binom{n}{k} \left(\frac{\lambda\tau}{n}\right)^k \left(1 - \frac{\lambda\tau}{n}\right)^{n-k}. \tag{20}$$

Additionally, we have

$$\lim_{n \to \infty} \left(1 - \frac{\lambda\tau}{n}\right)^{n-k} = e^{-\lambda\tau}, \tag{21}$$

and

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \underset{n \to \infty}{\approx} \frac{n^k}{k!}.$$

Hence, the PDF of a Poisson process is given by

$$P(k, \tau) = \lim_{n \to \infty} P(k, n) = \frac{(\lambda\tau)^k}{k!} e^{-\lambda\tau}. \tag{22}$$

Question: Does this PDF satisfy PDF requirements?

Definition 11. *A sequence of arrivals in continuous time is called a Poisson process with rate λ if*

1. *The number of arrivals in disjoint time intervals are independent.*

2. *The number of arrivals in an interval of length τ is given by the Poisson distribution*

$$P(k, \tau) = \frac{(\lambda\tau)^k}{k!} e^{-\lambda\tau}.$$

The expectation of the number of arrivals N_τ over an interval of length τ in a Poisson process is given by $\mathrm{E}[N_\tau] = \lambda\tau$. We also have $\mathrm{Var}[N_\tau] = \lambda\tau$.

Example: Let us assume that the arrivals of golfettes are a Poisson process with 1 golfette arrival every 10 minutes. What is the probability of 0, 1, 3 golfette arrivals within half an hour? We have $\lambda \tau = 1/10 * 30 = 3$

$$P(0, 30) = \frac{3^0}{0!} \times e^{-3} = e^{-3} = 0.05$$

$$P(1, 30) = \frac{3^1}{1!} \times e^{-3} = 3e^{-3} = 0.15$$

$$P(3, 30) = \frac{3^3}{3!} \times e^{-3} = e^{-3} = 0.22$$

Even when a problem makes no mention of Poisson processes, there are sometimes ways to pretend the rvs are coming from a Poisson process so that we can use convenient Poisson process properties.

3.1   Arrival times

Another interesting rv is the one describing the time to the N-th arrival. Let us denote it $T(n)$. Notice that $P(T(n) \leq t) = P(N(t) \geq n)$. Hence,

$$P(T(n) \leq t) = 1 - P(N(t) < n) = 1 - \sum_{k=1}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t} = \sum_{k=n}^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \tag{23}$$

Moreover, the PDF is obtained by deriving the CDF, which yields

$$f_{T_n}(t) = \frac{(\lambda t)^{k-1}}{(k-1)!} \lambda e^{-\lambda t}. \tag{24}$$

Alternatively, it can be proven using the definition of a density of probability

$$f_{T_n}(t) = \lim_{\delta \to 0} \frac{P(t < T_n < t + \delta)}{\delta}. \tag{25}$$

Notice that for $k = 1$, the time of the first arrival is given by the exponential distribution $f_{T_1}(t) = \lambda e^{-\lambda t}$.

Definition 12. *A positive rv X is memoryless if*

$$P(X > x_1 + x_2) = Pr(X > x_1)Pr(x > x_2), \quad \forall x_1, x_2 > 0. \tag{26}$$

Poisson processes are memoryless. The time to the next arrival is independent of the past. There are a couple of ways to picture this.

1. The first is the analogy with Bernoulli processes, which at the limit, define Poisson processes. In a Bernoulli process, the outcome of a coin flip does not impact the outcome of the next coin flip.

2. The second one if to take a real-world example. Consider phone calls that you get from strangers by mistake. The time at which the phone rings is independent from the previous ones. The process is memoryless.

Exercise 4. *Fishing*

3.2 Conditionning

Consider a Poisson process with conditionning on the total number of events in an interval?

Theorem 11. *Let $\{N(t), t > 0\}$ be a Poisson process with rate λ, and let $t_1 < t_2$. The conditional distribution of $N(t_1)$ given $N(t_2) = n$ is $N(t_1)|N(t_2) \sim \text{Bin}(n, t_1/t_2)$.*

.

3.3 Merging Poisson processes

If we take two independent Poisson processes and merge them, we get another Poisson process. This follows from the fact that the sum of independent Poissons is Poisson.

Theorem 12 (Merging). *Let $N_1(t), t > 0$ and $N_2(t), t > 0$ be independent Poisson processes with rates λ_1 and λ_2, respectively. The mergeed process $N(t) = N_1(t) + N_2(t)$ is a Poisson process with rate $\lambda_1 + \lambda_2$.*

Proof. We use the definition of a Poisson process.

- Arrivals in disjoint intervals are independent in the combined process because they are independent in the two individual processes, and the individual processes are independent of each other.

- The PMF of th merged process is given by

$$
\begin{aligned}
P(k, \tau) = \sum_{i=1}^{n} P_1(i, \tau) P_2(k - i, \tau) &= \sum_{i=1}^{n} \frac{(\lambda_1 \tau)^i}{i!} \lambda_1 e^{-\lambda_1 \tau} \frac{(\lambda_2 \tau)^{k-i}}{(k-i)!} \lambda_2 e^{-\lambda_2 \tau} \\
&= \sum_{i=1}^{n} \binom{k}{i} \frac{(\lambda_1 \tau)^i (\lambda_2 \tau)^{k-i}}{k!} e^{-(\lambda_1 + \lambda_2)\tau} \qquad (27) \\
&= \frac{((\lambda_1 + \lambda_2)\tau)^k}{k!} e^{-(\lambda_1 + \lambda_2)}.
\end{aligned}
$$

whereof the result. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Question: What is the probability that the arrival i from one process or the other?

Exercice: Suppose that student phones running out of charge are exponentially distributed with ditribution $\text{Exp}(\lambda)$ for each student. We decide not to leave the classroom until the last phone runs out of charge. When would we be able to leave the classroom (on average)?

An exponential rv is the 1st occurrence in a Poisson process. Consider a Poisson process that merges M processes, one for each student. The expected time until the 1st phone runs out of charge is $1/(M\lambda)$. We are now left with $M - 1$ students, and the same number of Poisson processes to merge. The expected time until the 2nd phone runs out of charge is $1/((M-1)\lambda)$. Thus, the expected time at which all phones have ran out of charge

$$
t = \frac{1}{M\lambda} + \frac{1}{(M-1)\lambda} + \ldots + \frac{1}{\lambda} = \frac{1}{\lambda} \sum_{i=1}^{M} \frac{1}{M}.
$$

3.4 Splitting Poisson processes

Consider a Poisson process. For each arrival, we independently flip a coin to decide whether it is a type-1 event or type-2 event, we end up with two independent Poisson processes. This is the converse of merging.

Theorem 13. *Let $\{N(t), t > 0\}$ be a Poisson process with rate λ, and classify each arrival in the process as a type-1 event with probability p and a type-2 event with probability $1 - p$, independently. Then the type-1 events form a Poisson process with rate λp, the type-2 events form a Poisson process with rate $\lambda(1 - p)$, and these two processes are independent.*

Proof. We use the definition of a Poisson process.

- Arrivals in disjoint intervals are independent in the splitted processes because they are independent in the individual parent process.

- The PMF of the type-1 splitted process is given by

$$P_1(k, \tau) = p^k \frac{(\lambda \tau)^k}{k!} \lambda e^{-\lambda \tau} \tag{28}$$

which is the PMF of a Poisson process of rate λp; whereof the result.

$\square$

4   Continuous-Time Markov chains

Definition 13. *A stochastic process $\{X(t) : t \geq 0\}$ with discrete state space S is called a continuous-time Markov chain (CTMC) if for all $t \geq 0$, $s \geq 0$, $i \in S$, and $j \in S$,*

$$P(X(s+t) = j \mid X(s) = i, \{X(u) : 0 \leq u < s\}) = P(X(s+t) = j \mid X(s) = i) = P_{ij}(t).$$

$P_{ij}(t)$ is the probability that the chain will be in state j, t time units from now, given it is in state i now. For each $t \geq 0$, there is a transition matrix

$$P(t) = (P_{ij}(t)).$$

4.1   Infinitesimal generator

The infinitesimal generator matrix $\boldsymbol{Q} = (q_{ij})_{1 \leq i,j \leq n}$ is given by

$$q_{ij} = \begin{cases} b_{ij}, & \text{if } j = i+1, \\ d_{ij}, & \text{if } j = i-1, \\ -b_{ij} - d_{ij}, & \text{if } i = j. \end{cases} \tag{29}$$

4.2   Forward Kolmogorov equation

On the other hand, the forward Kolmogorov differential equation describes the probability distribution of a state in time t keeping the initial point fixed. It decomposes the time interval $(0, t+s)$ into $(0, t)$ and $(t, t+s)$ by a so-called "last step analysis". We differentiate with respect to t then set $t = 0$, which yields

$$p'_{ij}(s) = \sum_{k=1}^{n} p_{ik}(s) p'_{kj}(0). \tag{30}$$

Using the infinitesimal generator, the Kolmogorow backward equation is then given by

$$\boldsymbol{P}'(s) = \boldsymbol{P}(s)\boldsymbol{Q} \tag{31}$$

4.3   Backward Kolmogorov equation

The backward Kolmogorov differential equation describes the transition probabilities in their dependence on the initial point i. Basically, it analyzes the time interval $(0, t+s)$ by the "first step analysis". It holds that
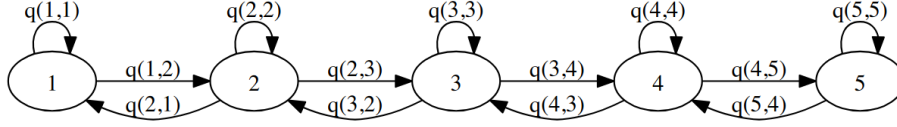
$$p_{ij}(h) = \delta_{ij} + h q_{ij} + o(h).$$

Thus,

$$p'_{ij}(0) = \frac{p_{ij}(h) - p_{ij}(0)}{h} = q_{ij}. \tag{32}$$

The CKE in continuous form is written as

$$p_{ij}(s+t) = \sum_{k=1}^{n} p_{ik}(s) p_{kj}(t). \tag{33}$$

We differentiate with respect to s then set $s = 0$, which yields

$$p'_{ij}(t) = \sum_{k=1}^{n} p'_{ik}(0)p_{kj}(t). \tag{34}$$

Using the infinitesimal generator, the Kolmogorow backward equation is then given by

$$\boldsymbol{P}'(t) = \boldsymbol{Q}\boldsymbol{P}(t) \tag{35}$$

The Kolmogogrov backward equation i a first order ODE whose solution is given by

$$P(t) = e^{\boldsymbol{Q}t} = \sum_{k=1}^{\infty} \frac{\boldsymbol{Q}^k t^k}{k!}. \tag{36}$$

4.4 Birth-death processes

A birth-death chain on states $\{1, 2, \ldots, M\}$ is a Markov chain with transition matrix $Q = (q_{ij})$ such that $q_{ij} > 0$ if $|i - j| = 1$ and $q_{ij} = 0$ if $|i - j| \geq 2$. This says it's possible to go one step to the left and possible to go one step to the right (except at boundaries) and possible to stay in the same state, but impossible to jump further in one step. The name stems from applications to the growth or decline of a population, where a step to the right is thought of as a birth and a step to the left is thought of as a death in the population. For example, the chain shown below is a birth-death chain if the labeled transitions have positive probabilities, except for the loops from a state to itself, which are allowed to have 0 probability. We will now show that any birth-death chain is reversible, and construct the stationary distribution. A birth death process is an irreducible aperiodic Markov chain. Hence, it has a unique stationary distribution $\boldsymbol{s}$ that is a reversible one. Reversibility can be shown using the fact that the frequency of moves forward is the same as that of moves backward because in order to move forward n times one needs to have moved backward $n - 1$ times at least. Thus long term frequencies of moves forward and backward between two nodes should be equal, i.e,

$$s_i p_{i(i-1)} = s_{i-1} p_{(i-1)i}$$

$$s_i = s_{i-1} \frac{p_{(i-1)i}}{p_{i(i-1)}} = s_{i-2} \frac{p_{(i-2)(i-1)}}{p_{(i-1)(i-2)}} \frac{p_{(i-1)i}}{p_{i(i-1)}}$$

Hence

$$s_i = s_0 \frac{\Pi_{k=1}^{i} p_{(k-1)k}}{\Pi_{k=1}^{i} p_{k(k-1)}}$$

We retrieve s_0 using the property $\sum_i s_i = 1$.

Suppose $p_{i(i+1)} = p$ and $p_{(i+1)i} = q$ for all $i \geq 1$, then

$$s_i = s_0(p/q)^i = s_0\rho^i,$$

Using $\sum_{i=0}^{n} s_i = 1$ yields $s_0 = 1 - \rho$, then $s_i = \rho^i(1 - \rho)$.

4.5 Holding times

For each state $i \geq 0$ we have a birth rate λ_i and a death rate μ_i. Whenever $X(t) = i$, independent of the past, the time until the next birth is a r.v. $X \sim \exp(\lambda_i)$ and, independently, the time until the next death is a r.v. $Y \sim \exp(\mu_i)$. Thus, the holding time rates are given by $a_i = \lambda_i + \mu_i$ because the time until the next transition (change of state) is given by the holding time $H_i = \min\{X, Y\} \sim \exp(\lambda_i + \mu_i)$. The idea here is that at any given time, the next birth is competing with the next death to be the next transition. (We always assume here that $\mu_0 = 0$ since there can be no deaths without a population.) This means that whenever $X(t) = i \geq 1$, the next transition will be a birth with probability

$$P_{i,i+1} = P(X < Y) = \frac{\lambda_i}{\lambda_i + \mu_i},$$

and a death with probability

$$P_{i,i-1} = P(Y < X) = \frac{\mu_i}{\lambda_i + \mu_i}.$$

Thus, the embedded chain for a B&D process is a simple random walk with state-dependent "up" and "down" probabilities.

When $\mu_i = 0$, $i \geq 0$, and $\lambda_i > 0$, $i \geq 0$, we call the process a *pure birth process*; the population continues to increase by one at each transition.

4.6 Explosion

Consider a pure birth process $\{X(t)\}$, $P_{i,i+1} = 1$, $i \geq 0$, in which $a_i = \lambda_i = 2^i$, $i \geq 0$. This process spends, on average, $E(H_i) = \frac{1}{\lambda_i} = 2^{-i}$ units of time in state i and then changes to state $i + 1$. Thus, it spends less and less time in each state, consequently jumping to the next state faster and faster as time goes on. Since $X(t) \to \infty$ as $t \to \infty$, we now explore how fast this happens.

Note that the chain will visit state i at time $H_0 + H_1 + \cdots + H_{i-1}$, the sum of the first i holding times. Thus, the chain will visit all of the states by time

$$T = \sum_{i=0}^{\infty} H_i.$$

Taking the expected value yields

$$E(T) = \sum_{i=0}^{\infty} 2^{-i} = 2 < \infty,$$

and we conclude that on average all states $i \geq 0$ have been visited by time $t = 2$, a finite amount of time! But this implies that with probability 1, all states will be visited in a finite amount of time; $P(T < \infty) = 1$. Consequently, with probability 1, $X(T + t) = \infty$, $t \geq 0$. This is an example of an explosive Markov chain: The number of transitions in a finite interval of time is infinite. A non-explosive CTMC is one for which the number of transitions in any finite interval of time is finite.

References

http://www.stat.yale.edu/Courses/1997-98/101/ranvar.htm (Definitions taken from Valerie J. Easton and John H. McColl's Statistics Glossary v1.1)

$file : ///home/benaceur/Documents/course_stochastic_processes/markov_cermics.pdf$