

# Math 525: Lecture 20

March 27, 2018

## 1 Limiting distribution

We have already seen many situations involving the *limiting distribution*

$$\alpha^\top = \lim_{n \rightarrow \infty} \mu^\top P^n \tag{1}$$

of a Markov chain, where  $\mu$  is the *initial distribution* (i.e.,  $\mathbb{P}(X_0 = i) = \mu_i$ ). To stress that  $\alpha$  is a function of  $\mu$ , we will write  $\alpha(\mu)$ .

**Example 1.1** (Gambler's ruin). We determined the probability of ruin by substituting

$$P = \begin{pmatrix} 1 & 0 & & & & & \\ 1/2 & 0 & 1/2 & & & & \\ & 1/2 & 0 & 1/2 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & 1/2 & 0 & 1/2 & \\ & & & & 0 & 1 & \end{pmatrix} \quad \text{and} \quad \mu = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

into (1). There, we found that

$$\alpha(\mu) = \begin{pmatrix} 1/2 \\ 0 \\ \vdots \\ 0 \\ 1/2 \end{pmatrix}.$$

More generally, it is natural to ask:

1. Is (1) always independent of the initial distribution  $\mu$ ?
2. Is (1) an equilibrium/stationary distribution? (i.e.,  $\alpha(\mu)^\top P = \alpha(\mu)^\top$ )

The answer to the first question is very easily seen to be no:

**Example 1.2.** Consider the transition matrix  $P = I$ . Then,  $P^n = I$  and hence

$$\alpha(\mu) = \lim_{n \rightarrow \infty} \mu^\top P^n = \lim_{n \rightarrow \infty} \mu^\top = \mu^\top.$$

Worse yet, the limit does not even have to exist:

**Example 1.3.** Consider

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \mu = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Note that  $P^{2n} = I$  and  $P^{2n+1} = P$ . Therefore,

$$\mu^\top P^{2n+1} = (0 \ 1) \quad \text{and} \quad \mu^\top P^{2n} = (1 \ 0)$$

so that  $\alpha(\mu)$  does not exist.

As for the second question, suppose the limit exists. Then, by continuity,

$$\alpha(\mu)^\top P = \left( \lim_{n \rightarrow \infty} \mu^\top P^n \right) P = \lim_{n \rightarrow \infty} \mu^\top P^{n+1} = \alpha(\mu)^\top.$$

For the remainder, we will try to understand when the limit is independent of the initial distribution. As in the gambler's ruin, the limiting distribution tells us a lot about the problem, and can even be considered as an approximation for  $\mu^\top P^n$  with  $n$  large. Moreover, we will see that in many of these cases, the limit can be computed efficiently.

## 2 Primitive matrices

**Definition 2.1.** Let  $A$  be a square matrix that is nonnegative (i.e.,  $A_{ij} \geq 0$ ). We say  $A$  is *primitive* if there exists a positive integer  $m$  such that  $A^m$  is positive (i.e.,  $(A^m)_{ij} > 0$ ).

**Lemma 2.2.** *If  $A$  is primitive, then  $A$  is irreducible.*

*Proof.* We prove this by contrapositive. Suppose  $A$  is reducible. By definition, we can find a permutation matrix  $K$  such that

$$KAK^\top = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}.$$

Let  $m$  be a positive integer. Then,

$$(KAK^\top)^m = KA^m K^\top = K \begin{pmatrix} B^m & X \\ 0 & D^m \end{pmatrix} K^\top$$

where  $X$  is some matrix (depending on  $B, C, D$ , and  $m$ ). The above implies that  $A$  is not primitive since  $A^m$  has some entries equal to zero.  $\square$

**Proposition 2.3.** *Let  $A$  be a nonnegative square matrix. Then,  $A$  is irreducible and aperiodic (i.e., has period equal to one) if and only if it is primitive.*

Before we give a proof, let's discuss some consequences. Given a matrix  $A$ , recall that

$$i \rightarrow j$$

means that we can find a walk

$$i = i_1 \dashrightarrow i_2 \dashrightarrow \cdots \dashrightarrow i_k = j$$

from  $i$  to  $j$ . We refer to the number of edges in a walk as its *length* (in the above example, the length is  $k - 1$ ). Recall also that the period of  $i$  is defined as

$$d(i) = \gcd(\{n \geq 1 : (A^n)_{ii} > 0\}).$$

Moreover,

$$(A^n)_{ii} > 0 \iff \text{there exists a walk of length } n \text{ from } i \text{ to itself.}$$

Therefore, we can equivalently define the period as

$$d(i) = \gcd(\{\text{length}(w) : w \text{ is a walk from } i \text{ to itself}\}).$$

This observation gives us a quick way to check if a matrix is primitive by checking the gcd of walks from a node to itself. One particularly useful consequence is given below.

**Corollary 2.4.** *Let  $A = (A_{ij})$  be a nonnegative and irreducible square matrix. If  $A_{ii} > 0$  for some  $i$ , then  $A$  is primitive.*

*Proof.* By the assumptions,  $d(i) = 1$  and hence  $A$  is aperiodic. Therefore, by Proposition 2.3,  $A$  is primitive.  $\square$

**Example 2.5.** The matrix

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

satisfies Corollary 2.4. Indeed, matrix multiplication reveals

$$P^4 = \begin{pmatrix} 0.5625 & 0.3125 & 0.1250 \\ 0.2500 & 0.2500 & 0.5000 \\ 0.6250 & 0.1250 & 0.2500 \end{pmatrix}.$$

Let's now return to our goal of proving Proposition 2.3. First, we need an intermediate result from number theory/algebra, which we state without proof.

**Lemma 2.6** (Semigroup lemma). *Any set of non-negative integers which is closed under addition and which has greatest common divisor 1 must contain all but finitely many of the non-negative integers.*

We will just prove the forward direction of Proposition 2.3:

*Proof of Proposition 2.3 ( $\Rightarrow$ ).* Let  $A$  be an irreducible and aperiodic matrix. For each row  $i$  of  $A$  let

$$J_i = \{n \geq 1: (A^n)_{ii} > 0\}.$$

Now, let  $n_1$  and  $n_2$  be elements of  $J_i$ . Then,

$$(A^{n_1+n_2})_{ii} = (A^{n_1}A^{n_2})_{ii} = \sum_k (A^{n_1})_{ik}(A^{n_2})_{ki} \geq (A^{n_1})_{ii}(A^{n_2})_{ii} > 0$$

and hence  $n_1 + n_2$  is also in  $J_i$ . That is,  $J_i$  is closed under addition. Moreover,  $\gcd(J_i) = 1$  by the presumed aperiodicity.

Applying lemma 2.6 to  $J_i$ , we can find  $M(i)$  such that for all  $n \geq M(i)$ , we have  $(A^n)_{ii} > 0$ . Since  $A$  is irreducible, we can find  $m(i, j)$  such that  $(A^{m(i, j)})_{ij} > 0$ . Therefore, for  $n \geq M(i)$ ,

$$(A^{n+m(i, j)})_{ij} = \sum_k (A^n)_{ik}(A^{m(i, j)})_{kj} \geq (A^n)_{ii}(A^{m(i, j)})_{ij} > 0.$$

Let  $M = \max_i \{M(i)\} + \max_{i, j} \{m(i, j)\}$ . Then,  $(A^M)_{ij} > 0$  for all  $i, j$  as desired.  $\square$

**Definition 2.7.** A Markov chain whose transition matrix is primitive is called *regular*.

**Proposition 2.8.** *Let  $P$  be the transition matrix of a regular Markov chain. Then, there exists a vector  $\alpha$  such that for any vector  $\mu$ ,*

$$\alpha^\top = \lim_{n \rightarrow \infty} \mu^\top P^n. \quad (2)$$

Moreover,

$$\alpha^\top P = \alpha^\top. \quad (3)$$

In fact,  $\alpha = c_1 v_1$  where  $v_1$  is a positive eigenvector corresponding to the eigenvalue  $\lambda = 1$  of  $P^\top$  and  $c_1$  is a normalizing constant which ensures that  $\alpha$  is a probability vector.

You are asked to prove this when  $P^\top$  admits a full set of linearly independent eigenvectors in assignment 8. The general case can also be proved using the Jordan decomposition of  $P^\top$ . Some observations:

- (2) states that the limiting distribution is independent of  $\mu$ .
- (3) states that the limiting distribution is an equilibrium of the Markov chain.
- $\alpha$  can be obtained by computing an eigenvector of  $P^\top$  associated with the eigenvalue  $\lambda = 1$ . For small matrices, you can do this by hand. In practice, there are various computational methods to do this, the simplest of which is the *power method*.

Of course, there are “irregular” Markov chains which still have limiting distributions:

**Example 2.9.** Note that

$$\mu^\top \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \mu_1 + \mu_2 \\ 0 \end{pmatrix}^\top$$

for any vector  $\mu$ .

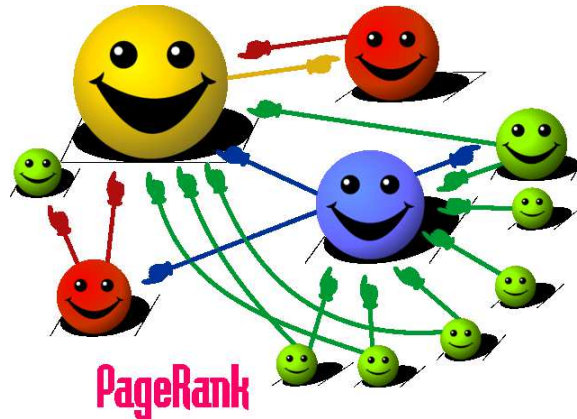


Figure 1: The “goodness” of each page is proportional to the number of links to that page.

### 3 Page Rank

We close by discussing an application of regular Markov chains: Page Rank. This is an algorithm used by Google to determine the “goodness” of a page so as to determine how to order the results of a search query.

Consider a web with  $N$  pages. Links between pages are recorded in a binary matrix  $G = (G_{ij})$ . That is,

$$G_{ij} = \begin{cases} 1 & \text{if there is a link from page } i \text{ to page } j \\ 0 & \text{otherwise.} \end{cases}$$

Letting  $e$  denote the column vector of ones in  $\mathbb{R}^N$ , note that

$$(Ge)_i = \text{number of outgoing links from } i.$$

We wish to model a *web surfer*. Suppose first that each page has at least one outgoing link. If the surfer is at page  $i$  at time  $n$ , then they choose a page from the set

$$\{j: G_{ij} = 1\}$$

to visit at time  $n + 1$  with uniform probability. Our assumption guarantees that the above is nonempty. The transition matrix corresponding to this Markov chain is

$$\tilde{P} = \text{diag}(Ge)^{-1}G$$

where

$$\text{diag}(x) = \begin{pmatrix} x_1 & & & \\ & x_2 & & \\ & & \ddots & \\ & & & x_N \end{pmatrix}$$

is the diagonal matrix obtained by placing the entries of the vector  $x$  on the diagonal. There are two problems with this preliminary model: (1) it’s not realistic and (2) it’s not a regular Markov chain.

We would like to incorporate the notion that the web surfer is *restless* and can, at any point in time, pick any page (not necessarily via an outgoing link) at random. Let

$$E = ee^T = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 1 & \cdots & 1 \\ 1 & 1 & \vdots & & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix}$$

where  $e$  is the vector of ones. Then, incorporating the restlessness, we arrive at the new transition matrix

$$P = (1 - \alpha) \text{diag}(Ge)^{-1}G + \alpha \frac{1}{m}E$$

where  $0 < \alpha \leq 1$ . The closer  $\alpha$  is to one, the “more restless” our web surfer. Since  $\frac{1}{m}E$  is a positive transition matrix,  $P$  is trivially a regular Markov chain.

*Remark 3.1.* The assumption that each page has at least one outgoing link is not particularly restrictive. Indeed, we can replace  $G$  by  $G'$ , which is defined by

$$G'_{ij} = \begin{cases} G_{ij} & \text{if } (Ge)_i > 0 \\ 1 & \text{otherwise.} \end{cases}$$

This corresponds to a surfer that, having encountered a page with no outgoing links, becomes restless and picks a page at random.

Since  $P$  is regular, the limiting distribution

$$\nu^T = \lim_{n \rightarrow \infty} \mu^T P^n$$

gives us an idea of the random surfers distribution after  $n$  steps, assuming  $n$  is large. Since  $P$  is regular, this does not depend on the initial distribution  $\mu$ . Now, let

$$\hat{\nu} = \begin{pmatrix} 1 & \nu_1 \\ 2 & \nu_2 \\ \vdots & \vdots \\ N & \nu_N \end{pmatrix}$$

and sort the rows of  $\hat{\nu}$  according to the second column:

**Example 3.2.** Suppose

$$\nu = \begin{pmatrix} 0.2 \\ 0.1 \\ 0.3 \\ 0.4 \\ 0.1 \end{pmatrix} \implies \hat{\nu} = \begin{pmatrix} 1 & 0.2 \\ 2 & 0.1 \\ 3 & 0.3 \\ 4 & 0.4 \\ 5 & 0.1 \end{pmatrix}.$$

After sorting,

$$\hat{\nu}_{\text{sorted}} = \begin{pmatrix} 4 & 0.4 \\ 3 & 0.3 \\ 1 & 0.2 \\ 2 & 0.1 \\ 5 & 0.1 \end{pmatrix}$$

and hence we can conclude that page 4 has the highest rank, page 3 the second highest, etc.