The Perron-Frobenius theorem, Google and Google-like things.

Positive matrices

The Perron-Frobenius theorem is a theorem about eigenvalues and eigenvectors of *positive* matrices. We say that a matrix,

$$P = (p_{ij})_{N \times N},$$

is positive if all $p_{ij} > 0$ (and nonnegative if all $p_{ij} \geq 0$).

Positive matrices

The Perron-Frobenius theorem is a theorem about eigenvalues and eigenvectors of *positive* matrices. We say that a matrix,

$$P = (p_{ij})_{N \times N},$$

is positive if all $p_{ij} > 0$ (and nonnegative if all $p_{ij} \geq 0$).

We say that P is *stochastic* if it is non negative and

$$\sum_i P_{ij} = 1$$

for all j.

Stochastic matrices

Think of a stochastic process (X_k) that at any time k can be in one of the states $1, 2, ...$ or N. Such a stochastic process is called a *Markov chain* if its probability distribution at a time $k + 1$ only depends on where it is at time k.

Stochastic matrices

Think of a stochastic process (X_k) that at any time k can be in one of the states $1, 2, ...$ or N. Such a stochastic process is called a *Markov chain* if its probability distribution at a time $k+1$ only depends on where it is at time k.

Let p_{ij} be the probability that the process at time $k+1$ is in state i, given that it is in state j the previous time k. Then $\sum_i p_{ij} = 1$ for all j, and all $p_{ij} \geq 0$. So, it is a stochastic matrix.

Stochastic matrices

Think of a stochastic process (X_k) that at any time k can be in one of the states $1, 2, ...$ or N. Such a stochastic process is called a *Markov chain* if its probability distribution at a time $k+1$ only depends on where it is at time k.

Let p_{ij} be the probability that the process at time $k+1$ is in state i, given that it is in state j the previous time k. Then $\sum_i p_{ij} = 1$ for all j, and all $p_{ij} \geq 0$. So, it is a stochastic matrix.

Say that the probability distribution at time k is given by $x = (x_1, ...x_N)$, so that the probility to be in the state i at time k is x_i.

Stochastic matrices

Think of a stochastic process (X_k) that at any time k can be in one of the states $1, 2, \ldots$ or N. Such a stochastic process is called a *Markov chain* if its probability distribution at a time $k+1$ only depends on where it is at time k.

Let p_{ij} be the probability that the process at time $k+1$ is in state i, given that it is in state j the previous time k. Then $\sum_i p_{ij} = 1$ for all j, and all $p_{ij} \geq 0$. So, it is a stochastic matrix.

Say that the probability distribution at time k is given by $x = (x_1, \ldots x_N)$, so that the probility to be in the state i at time k is x_i.

Then the probility distribution at time $k+1$ is given by the vector Px, i. e. the probability to be in state i is $\sum_j p_{ij} x_j$.

Stochastic matrices

Think of a stochastic process (X_k) that at any time k can be in one of the states $1, 2, ...$ or N. Such a stochastic process is called a *Markov chain* if its probability distribution at a time $k+1$ only depends on where it is at time k.

Let p_{ij} be the probability that the process at time $k+1$ is in state i, given that it is in state j the previous time k. Then $\sum_i p_{ij} = 1$ for all j, and all $p_{ij} \geq 0$. So, it is a stochastic matrix.

Say that the probability distribution at time k is given by $x = (x_1, ...x_N)$, so that the probility to be in the state i at time k is x_i.

Then the probility distribution at time $k+1$ is given by the vector Px, i. e. the probability to be in state i is $\sum_j p_{ij}x_j$.

The probability vector x is invariant if $Px = x$, i. e. x is an eigenvector of P with eigenvalue 1.

The Perron-Frobenius Theorem

We also say that a vector $x = (x_1, ...x_n)$ is positive if all $x_i > 0$ and nonnegative if all $x_i \geq 0$ (like probability distributions). We now state the first part of the Perron-Frobenius theorem.

Theorem

Let P be a positive matrix. Then P has exactly one positive eigenvector x^0. If λ_0 is its eigenvalue, then $\lambda_0 > 0$, and if P is stochastic, $\lambda_0 = 1$.

The Perron-Frobenius Theorem

We also say that a vector $x = (x_1, ...x_n)$ is positive if all $x_i > 0$ and nonnegative if all $x_i \geq 0$ (like probability distributions). We now state the first part of the Perron-Frobenius theorem.

Theorem

Let P be a positive matrix. Then P has exactly one positive eigenvector x^0. If λ_0 is its eigenvalue, then $\lambda_0 > 0$, and if P is stochastic, $\lambda_0 = 1$. If $\lambda > \lambda_0$, there is no nonnegative vector x with

$$Px \geq \lambda x.$$

Sketch of proof:

Let K_+ be the positive octant, i e the set of all non negative vectors. That P is positive implies that

$$P(K_+) \subset K_+.$$

Sketch of proof:

Let K_+ be the positive octant, i e the set of all non negative vectors. That P is positive implies that

$$P(K_+) \subset K_+.$$

We say that a subset K of K_+ is a *cone* if $x \in K$ implies $sx \in K$ for all $s > 0$.

Sketch of proof:

Let K_+ be the positive octant, i e the set of all non negative vectors. That P is positive implies that

$$P(K_+) \subset K_+.$$

We say that a subset K of K_+ is a *cone* if $x \in K$ implies $sx \in K$ for all $s > 0$. We say that K is an invariant cone if $PK \subset K$.

Sketch of proof:

Let K_+ be the positive octant, i e the set of all non negative vectors. That P is positive implies that

$$P(K_+) \subset K_+.$$

We say that a subset K of K_+ is a *cone* if $x \in K$ implies $sx \in K$ for all $s > 0$. We say that K is an invariant cone if $PK \subset K$. So, K_+ is a an invariant cone for P.

We are looking for an eigenvector of P in K_+. If such and eigenvector x_0 exists, it would define a minimal invariant cone, $\{sx^0\}$. So, we look for small invariant cones.

We are looking for an eigenvector of P in K_+. If such and eigenvector x_0 exists, it would define a minimal invariant cone, $\{sx^0\}$. So, we look for small invariant cones.

But, if P is positive $P(K_+)$ is strictly included in K_+. In fact, it is a relatively closed subcone.

We are looking for an eigenvector of P in K_+. If such and eigenvector x_0 exists, it would define a minimal invariant cone, $\{sx^0\}$. So, we look for small invariant cones.

But, if P is positive $P(K_+)$ is strictly included in K_+. In fact, it is a relatively closed subcone. Note that

$$P^{k+1}(K_+) \subset P^k(K_+).$$

We are looking for an eigenvector of P in K_+. If such and eigenvector x_0 exists, it would define a minimal invariant cone, $\{sx^0\}$. So, we look for small invariant cones.

But, if P is positive $P(K_+)$ is strictly included in K_+. In fact, it is a relatively closed subcone. Note that

$$P^{k+1}(K_+) \subset P^k(K_+).$$

Now look at

$$\bigcap_k P^k(K_+).$$

It is an invariant cone.

One can then show that, since $P^k(K_+)$ is strictly decreasing, its limit is just a half line, $L = \{sx^0; s > 0\}$. Since L is an invariant cone, x^0 is an eigenvector.

One can then show that, since $P^k(K_+)$ is strictly decreasing, its limit is just a half line, $L = \{sx^0; s > 0\}$. Since L is an invariant cone, x^0 is an eigenvector. This means that $Px^0 = \lambda_0 x^0$ and clearly $\lambda_0 > 0$.

Clearly, any other non negative eigenvector lies in L as well (why?), so it must be equal to (a multiple of) x^0. This proves the first part of the theorem.

One can then show that, since $P^k(K_+)$ is strictly decreasing, its limit is just a half line, $L = \{sx^0; s > 0\}$. Since L is an invariant cone, x^0 is an eigenvector. This means that $Px^0 = \lambda_0 x^0$ and clearly $\lambda_0 > 0$.

Clearly, any other non negative eigenvector lies in L as well (why?), so it must be equal to (a multiple of) x^0. This proves the first part of the theorem. The second, more technical, part said that:

If $\lambda > \lambda_0$, there is no nonnegative vector x with

$$Px \geq \lambda x.$$

Proof of the technical part

Let K be any closed invariant subcone of K_+ and take $x \in K$. Let $x^k = P^k x$.

Proof of the technical part

Let K be any closed invariant subcone of K_+ and take $x \in K$. Let $x^k = P^k x$. Then $x^k/|x^k| \in K$ for all k.

Proof of the technical part

Let K be any closed invariant subcone of K_+ and take $x \in K$. Let $x^k = P^k x$. Then $x^k/|x^k| \in K$ for all k. Moreover

$$x^k/|x^k| \in P^k(K_+),$$

so it must converge to $x^0/|x^0|$. So $L \subset K$ i. e. L is a subset of any closed invariant subcone.

Proof of the technical part

Let K be any closed invariant subcone of K_+ and take $x \in K$. Let $x^k = P^k x$. Then $x^k/|x^k| \in K$ for all k. Moreover

$$x^k/|x^k| \in P^k(K_+),$$

so it must converge to $x^0/|x^0|$. So $L \subset K$ i. e. L is a subset of any closed invariant subcone. Take $\lambda > 0$ and look at

$$\{x \geq 0; Px \geq \lambda x\}.$$

It is a closed invariant subcone of K_+. Therefore, if it is not empty, it contains L and therefore x_0. So

$$\lambda_0 x_0 = Px_0 \geq \lambda x_0.$$

Hence, $\lambda \leq \lambda_0$.

The next part of the Perron-Frobenius theorem is

Theorem

All (complex) eigenvalues of P satisfy $|\lambda| \leq \lambda_0$. Equality holds only if the eigenvector is positive.

Bevis.

If $z = (z_1, ...z_n)$ is a complex eigenvector, then

$$\sum_j p_{ij} z_j = \lambda z_i.$$

The triangle inequality gives,

$$|\lambda||z_i| \leq \sum p_{ij} |z_j|.$$

By the previous part, $\lambda \leq \lambda_0$ Equality can hold only if all z_j have the same argument, so z can be taken positive. $\square$

Finally

Theorem

If P is positive and stochastic, $\lambda_0 = 1$.

Finally

Theorem

If P is positive and stochastic, $\lambda_0 = 1$.

Bevis.

Let x be the eigenvector with $\sum x_i = 1$,

$$\lambda_0 x_i = \sum p_{ij} x_j.$$

Finally

Theorem

If P is positive and stochastic, $\lambda_0 = 1$.

Bevis.

Let x be the eigenvector with $\sum x_i = 1$,

$$\lambda_0 x_i = \sum p_{ij} x_j.$$

Sum over i:

$$\lambda_0 = \lambda_0 \sum x_i = \sum_{ij} p_{ij} x_j = 1.$$

$\square$

If P is stochastic,

$$x_0 = \lim_{k\to\infty} P^k x',$$

for any $x' \neq 0$.

If P is stochastic,

$$x_0 = \lim_{k \to \infty} P^k x',$$

for any $x' \neq 0$. This is at least roughly clear since all the other eigenvalues have eigenvalues smaller than 1.

Application to Google

There is also a version of the theorem for non negative matrices, but I will not go into that. We will now discuss an application of the theorem.

Application to Google

There is also a version of the theorem for non negative matrices, but I will not go into that. We will now discuss an application of the theorem. This is the *Page rank algorithm* of Google. It deals with the problem:

How do you rank the popularity of a webpage?

Application to Google

There is also a version of the theorem for non negative matrices, but I will not go into that. We will now discuss an application of the theorem. This is the *Page rank algorithm* of Google. It deals with the problem:

How do you rank the popularity of a webpage? It is a special case of a more general problem: How do you rank (the popularity of) anything?

Application to Google

There is also a version of the theorem for non negative matrices, but I will not go into that. We will now discuss an application of the theorem. This is the *Page rank algorithm* of Google. It deals with the problem:

How do you rank the popularity of a webpage? It is a special case of a more general problem: How do you rank (the popularity of) anything? Researchers, football teams, people?

Application to Google

There is also a version of the theorem for non negative matrices, but I will not go into that. We will now discuss an application of the theorem. This is the *Page rank algorithm* of Google. It deals with the problem:

How do you rank the popularity of a webpage? It is a special case of a more general problem: How do you rank (the popularity of) anything? Researchers, football teams, people? We start with the popularity of people.

A person is popular if (s)he has many friends.

Application to Google

There is also a version of the theorem for non negative matrices, but I will not go into that. We will now discuss an application of the theorem. This is the *Page rank algorithm* of Google. It deals with the problem:

How do you rank the popularity of a webpage? It is a special case of a more general problem: How do you rank (the popularity of) anything? Researchers, football teams, people? We start with the popularity of people.

A person is popular if (s)he has many friends. But, not all friends are equal; true popularity consists in having many *popular* friends.

Application to Google

There is also a version of the theorem for non negative matrices, but I will not go into that. We will now discuss an application of the theorem. This is the *Page rank algorithm* of Google. It deals with the problem:

How do you rank the popularity of a webpage? It is a special case of a more general problem: How do you rank (the popularity of) anything? Researchers, football teams, people? We start with the popularity of people.

A person is popular if (s)he has many friends. But, not all friends are equal; true popularity consists in having many *popular* friends. This sounds like a circular definition, but can be made precise. Label all the people from 1 to N, and say the popularity of individual j is x_j. Say j is friends with N_j other people (this may be a non symmetrical relation).

Application to Google

There is also a version of the theorem for non negative matrices, but I will not go into that. We will now discuss an application of the theorem. This is the *Page rank algorithm* of Google. It deals with the problem:

How do you rank the popularity of a webpage? It is a special case of a more general problem: How do you rank (the popularity of) anything? Researchers, football teams, people? We start with the popularity of people.

A person is popular if (s)he has many friends. But, not all friends are equal; true popularity consists in having many *popular* friends. This sounds like a circular definition, but can be made precise. Label all the people from 1 to N, and say the popularity of individual j is x_j. Say j is friends with N_j other people (this may be a non symmetrical relation). Then we say that each of them inherits a fraction of j:s popularity

$$(1/N_j)x_j.$$

The total popularity of i becomes

$$x_i = \sum_j p_{ij} x_j,$$

where $p_{ij} = (1/N_j)$ if j is friends with i and 0 if not. This is a stochastic matrix since

$$\sum_i p_{ij} = (1/N_j)\#\{i; j \text{ is friends with } i\} = 1.$$

The total popularity of i becomes

$$x_i = \sum_j p_{ij} x_j,$$

where $p_{ij} = (1/N_j)$ if j is friends with i and 0 if not. This is a stochastic matrix since

$$\sum_i p_{ij} = (1/N_j)\#\{i; j \text{ is friends with } i\} = 1.$$

(There is a problem that P is not strictly positive that we will come to later.)

The total popularity of i becomes

$$x_i = \sum_j p_{ij} x_j,$$

where $p_{ij} = (1/N_j)$ if j is friends with i and 0 if not. This is a stochastic matrix since

$$\sum_i p_{ij} = (1/N_j)\#\{i; j \text{ is friends with } i\} = 1.$$

(There is a problem that P is not strictly positive that we will come to later.)
By Perron-Frobenius, it has an essentially unique solution and this is the ranking.

The total popularity of i becomes

$$x_i = \sum_j p_{ij} x_j,$$

where $p_{ij} = (1/N_j)$ if j is friends with i and 0 if not. This is a stochastic matrix since

$$\sum_i p_{ij} = (1/N_j)\#\{i; j \text{ is friends with } i\} = 1.$$

(There is a problem that P is not strictly positive that we will come to later.)
By Perron-Frobenius, it has an essentially unique solution and this is the ranking.

The same argument applies to internet pages. Then we replace people by internet pages and the friendship relation is : *there is a link from j to i*. In this case, we can interpret the ranking differently.

Imagine a person that is surfing the internet. At any page, (s)he clicks on a random (uniformly distributed) link from that page to the next.

Imagine a person that is surfing the internet. At any page, (s)he clicks on a random (uniformly distributed) link from that page to the next.

Given that at one time he is in position j, the probability that in the next time he is in position i is $1/N_j$ or 0, i. e. p_{ij}. Then we get a Markov chain and

$$x^0 = \lim_{k \to \infty} P^k x',$$

where $x' \neq 0$ is an arbitrary starting point.

Imagine a person that is surfing the internet. At any page, (s)he clicks on a random (uniformly distributed) link from that page to the next.

Given that at one time he is in position j, the probability that in the next time he is in position i is $1/N_j$ or 0, i. e. p_{ij}. Then we get a Markov chain and

$$x^0 = \lim_{k\to\infty} P^k x',$$

where $x' \neq 0$ is an arbitrary starting point. So x_i^0 is the probability that our surfer is at page i after a long time, or the proportion of the time he is there.

In practice, one makes P strictly positive by replacing it by

$$P^\beta := (1 - \beta)P + \beta Q,$$

where Q is some arbitrary positive stochastic matrix (like $q_{ij} = 1/N$ for all i, j.)

In practice, one makes P strictly positive by replacing it by

$$P^{\beta} := (1-\beta)P + \beta Q,$$

where Q is some arbitrary positive stochastic matrix (like $q_{ij} = 1/N$ for all i, j.)
Google updates its page rank monthly. This involves computing an eigenvector of an $N \times N$ where N is the number of web pages in the world.