


**$\sum p^{(n)} = \infty \leftrightarrow \sum f^{(n)} = 1$:**

We shall now prove the above statement thus completing the discussion of Random Walks. This involves interesting ideas which are valid in more generality.

**Renewal Equation:**

$$p^{(n)} = \sum_{k=0}^{n} f^{(k)} p^{(n-k)}; \qquad n \geq 1.$$

Recall $p^{(n)} = P(X_n = 0)$. The event $A = (X_n = 0)$ can be expressed as disjoint union of events

$$A = \bigcup_{1}^{n} A_k; \quad A_k = (X_i \neq 0, i < k; \quad X_k = 0; X_n = 0).$$

If you take any sample point for which $X_n(\omega) = 0$ then there must be a first $k \leq n$ such that $X_k(\omega) = 0$ and hence the above equation holds. By rules about conditional probability,

$$P(A_k) = P(X_i \neq 0, i < k;\ X_k = 0) P(X_n = 0 | X_i \neq 0, i < k;\ X_k = 0).$$
$$= f^{(k)}\ P(X_n = 0 | X_i \neq 0, i < k;\ X_k = 0).$$

Note that under the given condition, irrespective of whatever be $X_i$ for $i < k$; $(X_n = 0)$ holds when $(X_k = 0)$ iff in the $(n-k)$ days having now started at zero, you end at zero. Thus

$$P(A_k) = f^{(k)}\ p^{(n-k)}.$$

Thus, noting $f^{(0)} = 0$;

$$p^{(n)} = \sum_{k=1}^{n} f^{(k)} p^{(n-k)} = \sum_{k=0}^{n} f^{(k)} p^{(n-k)}$$

**generating functions:**

Let us define

$$P(s) = p^{(0)} + p^{(1)}s + p^{(2)}s^2 + \cdots + p^{(k)}s^k + \cdots$$
$$F(s) = f^{(0)} + f^{(1)}s + f^{(2)}s^2 + \cdots + f^{(k)}s^k + \cdots$$

which are called generating functions of the numbers $\{p^{(n)} : n \geq 0\}$ and the numbers $\{f^{(n)} : n \geq 0\}$ respectively. These functions are defined at least for

$$0 \leq s < 1$$

of course $f^{(n)}$ being probabilities of disjoint events, $F$ is defined for $s = 1$ as well.

Observe that, irrespective of whether $\sum p^{(n)}$ is finite or not we have,

$$\lim_{s \uparrow 1} P(s) = \sum_n p^{(n)}; \qquad \lim_{s \uparrow 1} F(s) = \sum_n f^{(n)}.$$

**Cauchy product of series:**

Recall that if $\sum a_n$ and $\sum b_n$ are two series of numbers then we can define another series, Cauchy product, as $\sum c_n$ where

$$c_n = a_0 b_n + a_1 b_{n-1} + a_2 b_{n-2} + \cdots + a_n b_0$$

Here then is Cauchy's theorem: If $\sum a_n = A$ and $\sum b_n = B$ and if at least one of these series is absolutely convergent, then the series $\sum c_n$ converges and converges to the product $AB$.

If you now fix an $s$ $(0 \leq s < 1)$ and consider the two series defining $F(s)$ and $P(s)$. Note they these being of positive terms they are absolutely convergent and thus we see by Cauchy's theorem and renewal equation,

$$F(s)P(s) = P(s) - 1.$$

We have here used that renewal equation is valid for $n \geq 1$, not for $n = 0$.

$$f^{(0)}p^{(0)} = 0; \qquad p^{(0)} = 1.$$

Thus

$$P(s) = \frac{1}{1 - F(s)}; \quad 0 \le s < 1. \qquad (\spadesuit)$$

Note that $F(s) < 1$ for $0 \le s < 1$, so the above makes sense. Moreover, if $\sum f^{(n)} = 1$ then $\lim_{s \uparrow 1} F(s) = 1$ and $(\spadesuit)$ shows that

$$\lim_{s \uparrow 1} P(s) = \infty; \qquad \text{or} \qquad \sum p^{(n)} = \infty.$$

On the other hand if $\sum f^{(n)} = c < 1$ then $\lim_{s \uparrow 1} F(s) = c$ and $(\spadesuit)$ shows that

$$\lim_{s \uparrow 1} P(s) = \frac{1}{1 - c} < \infty; \qquad \text{or} \qquad \sum p^{(n)} < \infty.$$

This completes proof of the theorem.

**Two state chain:**

Before discussing genera,l theory of Markov chains, let us discuss one more example. A chain that has two states $A$ and $B$. Here is the transition matrix.

$$\begin{array}{ccc} & A & B \\ A & 1-a & a \\ B & b & 1-b \end{array}$$

You can think of this as describing state of a machine. $A$ is 'on' and $B$ is 'off'. if the machine is on today, chances that it is off tomorrow are $a$; if it is off today then chances that it is on tomorrow is $b$.

**(1) Both $a$ and $b$ are zero**

Then the transition matrix $P$ is

$$\begin{array}{ccc} & A & B \\ A & 1 & 0 \\ B & 0 & 1 \end{array}$$

the identity matrix. Thus the chain stays in the starting state; there is no motion at all. $P^n$ is identity matrix for all $n$.

if you randomize initial state, say start on day zero by tossing a coin: Thus $X_0 = A$ with probability $p$ and $X_0 = B$ with probability $(1 - p)$ then distribution of $X_n$ remains so for ever.

**(2) Both $a$ and $b$ are one:**

The transition matrix $P$ now is

$$\begin{array}{ccc} & A & B \\ A & 0 & 1 \\ B & 1 & 0 \end{array}$$

In this case all even $P^n$ are identity matrix and all odd $P^n$ are the above matrix. Every day the state changes from what it was. The sequence of matrices $\{P^n\}$ does not converge.

if you randomize initial state, say start on day zero by tossing a coin: Thus $X_0 = A$ with probability $p$ and $X_0 = B$ with probability $(1-p)$ then distribution of $X_n$ remains so for all odd $n$ and it would be $(1-p, p)$ for all even $n$.

In particular, if you select $A$ or $B$ at random on initial day, then distribution of $X_n$ remains so for ever. In other words, if someone is keeping track of 'not the state' but only the 'distribution of the state' each day; then she will find it unchanging, that is, it remains stationary.

However, you can not say that this is equilibrium distribution because, if you started from any fixed state, you will not reach this distribution at all. There is no steady state.

**(3) they are zero-one. $a = 0$, $b = 1$:**

The transition matrix $P$ is now

$$\begin{array}{ccc} & A & B \\ A & 1 & 0 \\ B & 1 & 0 \end{array}$$

In this chain, if you start at $A$, you stay there like in first case. But if you start in $B$, you do not stay there, you immediately move to $A$ and of course form day one onwards you are in $A$. Thus no matter where you start, you are absorbed at $A$.

The other case $a = 1$ and $b = 0$ is similar.

**(4) one is zero and other not zero/one.** $a = 0$**,** $0 < b < 1$**:**

The transition matrix $P$ now is

$$\begin{array}{ccc} & A & B \\ A & 1 & 0 \\ B & b & 1-b \end{array}$$

if you start at $A$, you stay there. If you start at $B$ then you eventually (sooner or later, at a finite time) end up in $A$. We can not give any bound on how long we need to wait. We end up in $A$ on 'day $k$ for the first time' only by staying first $(k-1)$ days at $B$ and then moving to $A$. Thus if $T$ is the time of entering $A$ then $T$ takes values $\{1, 2, 3, \cdots\}$ and

$$P(T = k) = b(1-b)^{k-1}; \qquad k = 1, 2, 3, \cdots .$$

This is exactly like waiting time for heads in coin tossing. waiting time for entering $A$ has thus geometric distribution.

**(5) Both non-zero,non-one:** $0 < a < 1$ **and** $0 < b < 1$**:**

In this case the transition matrix $P$ is

$$\begin{array}{ccc} & A & B \\ A & 1-a & a \\ B & b & 1-b \end{array}$$

with all entries strictly between zero and one.

In this case if you consider $P^n$ we can show (shall do later) that $P^n$ converges entry wise to the matrix

$$\begin{array}{ccc} & A & B \\ A & \frac{b}{a+b} & \frac{a}{a+b} \\ B & \frac{b}{a+b} & \frac{a}{a+b} \end{array}$$

In other words if we start from either $A$ or $B$, we see

$$P(X_n = A) \to \frac{b}{a+b}; \quad P(X_n = B) \to \frac{a}{a+b}.$$

If you consider the probability vector

$$\pi = \left( \frac{b}{a+b}, \frac{a}{a+b} \right)$$

then

$$\pi P = P.$$

Thus $\pi$ is stationary probability. That is, if the initial distribution is $\pi$ then it remains so for ever; every $X_n$ has the same distribution. Something more is also true. It is steady state or equilibrium distribution as well: No matter where you start, the distribution of state at time $n$ converges to this $\pi$, as $n \to \infty$. Even if you randomized and started by tossing coin, the same happens, the distribution of your state on $n$-th day converges to $\pi$.

**Markov chains:**

Let us do some generalities now before returning to examples again. First recall definitions.

There is a countable set $S$ called **state space.** There is probability vector $\{\mu(i) : i \in S\}$ called **initial distribution.** There is a matrix

$$((P)) = (P_{ij} : i, j \in S\}$$

called **transition matrix** . This is a stochastic matrix, that is,

$$P_{ij} \geq 0, \ \forall \ \ i, j \in S; \qquad\qquad \sum_j P_{ij} = 1, \ \forall \ \ i \in S.$$

Idea is: whenever you find yourself in state $i$, you move to $j$ with probability $P_{ij}$.

A **markov chain** with above data is a sequence of random variables $(X_n : n \geq 0)$ taking values in $S$ such that

$$P(X_0 = i) = \mu(i); \quad \forall i \in S,$$

$$P(X_n = i_n | X_0 = i_0, X_1 = i_1, \cdots, X_{n-1} = i_{n-1})$$
$$= P(X_n = i_n | X_{n-1} = i_{n-1}) \ \ = \ \ p_{i_n i_{n-1}}$$

for all states $\{i_0, i_1, \cdots, i_n\}$ and for all $n \geq 1$.

This is called **Markov property**. On any day, irrespective of how you reached the state, say $i$, on that day you move to a state $j$ with probability $P_{ij}$ and continue.

Thus the three data items $(S, \mu, P)$ tell you where you are moving; how you start the motion; how you proceed.

There are two comments in order.

(1) So far we understood random variables as only real valued functions on a sample space. Now we are talking about functions taking values in $S$. Thus the definition of random variable is now more general.

(2) So far we understood sample space as a countable set with probability attached to each sample point. Where is the probability space now? well, you *can not* construct such a sequence of random variables on a countable set. You need a bigger platform to define these random variables. But there is no need for us to enter the discussion regarding uncountable sample spaces. Shall do so later. You just believe that there is a sequence of random variables satisfying the above condition.

Our objective is to calculate probabilities and answer questions. You can calculate the distributions of any finitely many of these random variables using the Markov property, no need to know the probability space. Here is how.

$$P(X_0 = i) = \mu(i)$$

we are told so.

$$P(X_0 = i, X_1 = j) = P(X_0 = i)P(X_1 = j|X_0 = i) = \mu(i)P_{ij}.$$

In particular, with above data,

$$P(X_1 = j) = \sum_i \mu(i)P_{ij} = j\text{-th entry of } \mu P.$$

$$P(X_0 = i, X_2 = j, X_3 = k) =$$
$$P(X_0 = i)P(X_1 = j|X_0 = i)P(X_3 = k|X_0 = i, X_1 = j)$$
$$= \mu(i)P_{ij}P_{jk}.$$

Here we used the Markov property for the third term. In particular, with above data

$$P(X_2 = k) = \sum_{i,j} \mu(i)P_{ij}P_{jk} \ = \ k\text{-th entry of } \mu P^2.$$

The distribution of $X_n$ is the vector $\mu P^n$.

Here is how the data is reflected in our notation.

$$P_\mu(X_n = k) \; = \; k\text{-th entry of } \; \mu P^n.$$

When the initial distribution puts mass 1 at a state $i$; that is you start at state $i$ we denote the probability by $P_i$. Thus

$$P_i(X_n = k) \; = \; (i, k)\text{-th entry of } \; P^n$$

$$= P_{ik}^{(n)}$$

Some times, instead of $P_i(X_n = k)$ we use the notation $P(X_n = k | X_0 = i)$. Both mean the same.

In other words, $i$-th row of $P^n$ gives you the distribution of your state on day $n$ if you started at state $i$ initially, just as $i$-th row gives you the distribution of your state on day 1 if you started at $i$ on day zero.

Obviously

$$P_\mu(X_n = k) = \sum_i \mu(i) P_i(X_n = k)$$

**Chapman-Kolmogorov Equations:**

$$P_{ij}^{(m+n)} = \sum_k P_{ik}^{(m)} P_{kj}^{(n)}.$$

stands to reason because we have for matrix multiplication

$$P^{m+n} = P^m P^n.$$

Equivalently the above equations can be stated as

$$P_i(X_{m+n} = j) = \sum_k P_i(X_m = k) P_k(X_n = j)$$

This is because of the identifications made above for these matrix elements. This has the following interpretation: if you started at $i$ and want to be at $j$ on day $(m+n)$ then you should be at some state $k$ on day $m$ and *starting from that state* $k$ you should reach $j$ on the $n$-th day from then.

Proof is also simple, postponing attention to some subtleties, here is the proof

$$P_i(X_{m+n} = j) = \sum_k P_i(X_{m+n} = j; X_m = k)$$

$$= \sum_k P_i(X_m = k) P_i(X_{m+n} = j | X_m = k)$$

The last term is simply probability of $(X_{m+n} = j)$ given $(X_m = k)$ because, by Markovian nature the initial state does not matter. But this is same as starting from $k$ now, moving to $j$ on day $n$ from now, that is $P_k(X_n = j)$. [shall attend to rigour later].

**communicating classes:**

Imagine state space being the set of all chairs in this room and then a law of motion is prescribed. Suppose the law is such that you move only to a chair in that row. Then to understand the Chain, you can try to look at the motion on each row and understand them. In other words, you can reduce the state space to appropriate subsets and understand the chair on each part separately. This is what we do now. In a sense, we now understand the topology of the state space that arises from the motion.

For two states $i, j$ we say $i \rightsquigarrow j$ (read $i$ **leads** to $j$) if for some $n \geq 1$ we have $P_{ij}^{(n)} > 0$. Thus starting from $i$ there is a chance of reaching $j$ on *some* day.

Say $i \leftrightsquigarrow j$ (read $i$ **communicates** with $j$) if $i \rightsquigarrow j$ and $j \rightsquigarrow i$.

Here rate some simple facts.

(1) If $i \rightsquigarrow j$ and $j \rightsquigarrow k$ then $i \rightsquigarrow k$.

Indeed $P_{ij}^{(m)} > 0$ and $P_{jk}^{(n)} > 0$ then Chapman-Kolmogorov equations tell you

$$P_{ik}^{(m+n)} \quad = \quad \sum_l P_{il}^{(m)} P_{lk}^{(n)} \quad \geq \quad P_{ij}^{(m)} P_{jk}^{(n)} \quad > \quad 0$$

(2) If $i \leftrightsquigarrow j$, then $j \leftrightsquigarrow i$.

Indeed to show this we need to show $i \rightsquigarrow j$ and $j \rightsquigarrow i$ and both follow from hypothesis.

(3) If $i \leftrightsquigarrow j$ and $j \leftrightsquigarrow k$ then $i \leftrightsquigarrow k$.

Indeed, we need to show that $i \rightsquigarrow k$ and $k \rightsquigarrow i$ and both follow from hypothesis and (1) above.

For each state $i$, let us denote by $C(i)$ the collection of states: $i$ and all states that communicate with $i$. This is called **communicating class of** $i$. Note that $i$ need not lead to $i$ and so need not communicate with itself.

Hence we have defined $C(i)$ as those states which communicate with $i$ and also the state $i$.

(4) For $i \neq j$; either $C(i) = C(j)$ or $C(i) \cap C(j) = \emptyset$.

Indeed if we temporarily define $i \sim j$ if either $i = j$ or $i \leftrightsquigarrow j$ then we see that $C(i)$ consists of all states $j \sim i$. Note that defining clause already tells that $i \sim i$. Properties (2) and (3) tell you that (i) $i \sim j$ implies $j \sim i$ and (ii) $i \sim j; j \sim k$ imply $i \sim k$. In other words $\sim$ is an equivalence relation and $C(i)$ are just nothing but $\sim$ equivalence classes. And so the statement follows.

Thus the state space is partitioned into subsets.

In the two state chain with state space $\{A, B\}$ and transition matrix consisting of each row $(1, 0)$ we see that the classes are $C(A) = \{A\}$ and $C(B) = \{B\}$. You can restrict the chain to the class $C(A)$ because if you start here you stay here. But note that it makes no sense to say 'restrict the chain to $C(B)$' because with our rule of motion if you start in $B$, you are moving outside $C(B)$.

Thus you can restrict chain to those communicating classes which do not lead outside. What about classes which lead outside? Shall see later.

A chain where there is only one communicating class is called **irreducible**. In other words if each $C(i)$ is the full state space $S$ then it is irreducible. Thus irreducible means for any two states $i$ and $j$ we have $i \leftrightsquigarrow j$ or equivalently, for any two states $i$ and $j$ we have $i \rightsquigarrow j$.

**Period:**

In the two state chain with transition matrix $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, you see that starting from a state you can return to that state only on even days. Similarly if you consider the random Walk on $Z$, then starting from zero, you can return to zero only on even days. In other words all return times are multiples of 2. This is what we capture now.

Let us consider a state $i$ such that $i \rightsquigarrow i$. That is for some $n \geq 1$ we have $P_{ii}^{(n)} > 0$. We define

$$d(i) = g.c.d\ \{n \geq 1 : P_{ii}^{(n)} > 0\}$$

This is called **period** of the state $i$. The definition is sensible. Firstly, the set above is non-empty. Secondly, if you take any one number $k$ in the set above then you need to search for the greatest common divisor of the set only among the finitely many numbers $\{1, 2, \cdots, k\}$. Since this finite set has a maximum, the gcd of the infinite set above is well defined.

Moreover this gcd is actually attained for a finite subset. That is there is a finite subset $F \subset \{n \geq 1 : P_{ii}^{(n)} > 0\} = A$ (say) such that $d(i) = gcd(F)$. This can be easily seen as follows. Take any element $n_1 \in A$ Set $g_1 = n_1$. if it divides all elements of $A$ we are done. If not take $n_2 \in A$ not divisible by $n_1$. set $g_2 = gcd\{n_1, n_2\}$. If it divides all elements of $A$ we are done. Otherwise, take $n_3 \in A$ not divisible by $g_2$ and make $g_3$ etc. This process stops at a finite stage because, by choice, $g_1 > g_2 > g_3 > \cdots$ and there is no infinite strictly decreasing sequence of non-negative integers.

We say that a state $i$ is **aperiodic** if $d(i) = 1$. A **chain is aperiodic** if every state is aperiodic. More on this later.

Here is one of the few basic theorems concerning finite state Markov chains.

■ **Theorem:** FINITE STATE SPACE.

Consider a markov chain with finite state space $S$ and transition matrix $P$. Assume that the chain is irreducible and aperiodic. Then there is a matrix $\Pi$ such that the following happen.

(1) $P^n \to \Pi$ as $n \to \infty$ (entrywise convergence).

(2) $\Pi$ has identical rows. That is there is one vector $\pi = (\pi_i : i \in S)$ such that each row of $\Pi$ consists of $\pi$.

(3) $\pi$ is a probability vector.

(4) $\pi$ satisfies $\pi P = \pi$.

(5) If $\eta$ is a probability vector such that $\eta P = \eta$ then we must have $\eta = \pi$.■

So what does the above theorem say? Let $(X_n : n \geq 0)$ be the Markov chain, say starting at one state $i$, fixed for now. We know that the $i$-th row of $P^n$ is the distribution of our state on day $n$. For each $j \in S$, we have

$$P_{ij}^{(n)} \to \pi_j; \quad i.e. \quad P(X_n = j | X_0 = i) \to \pi_j.$$

Thus the chain is reaching a steady state, in terms of probability distribution. Thus if the chain is running for a long time, the chances of finding it in state $j$ equals $\pi_j$.

That the rows of $\Pi$ all equal $\pi$ tells us that initial state does not matter. In other words, we need not ask as to where the chain started. No matter where it started, the chances of being in state $j$ equals $\pi_j$ after a long time elapsed. This remains so even if you randomise your state. This is because,

$$\begin{aligned} P_\mu(X_n = j) &= \sum_i \mu(i) P_i(X_n = j) \rightarrow \\ &= \sum_i \mu(i)\pi_j = \pi_j \end{aligned}$$

In other words, whatever be the mechanism of starting, whether a fixed state or at a state chosen according to some probability, we have the chances of seeing the chain in state $j$ equals $\pi_j$.

This is called steady state distribution or equilibrium distribution.

The equation $\pi = \pi P$ tells that if $X_0 \sim \pi$ then $X_1 \sim \pi$ and in fact for all $n$ we have $X_n \sim \pi$. In other words if you are running the chain with initial state having distribution $\pi$, then some one who is only keeping track of the distribution of the state for each day (and not keeping track of the exact state of the day) will feel that the chain is stationary, not moving at all.

For a markov chain with transition matrix $P$, a distribution $\pi$ (that is, a probability on the state space) is called **stationary distribution** if $\pi P = \pi$.

**Equilibrium distribution or steady-state distribution** is the limiting distribution of $X_n$, if there is any. It is quite likely that when you start from one initial state, the distribution of $X_n$ may converge to some thing and on the other hand when you start from another state the distribution of $(X_n)$ may still converge but to a different distribution.

what the above theorem says is that for an irreducible aperiodic chain, there is only one stationary distribution and *it is also* equilibrium distribution, no matter how you start.

We shall prove the following fact next time.

■ **Theorem:** A finite-state markov chain with transition matrix $P$ is irreducible and aperiodic iff there is an $n$ such that all entries of $P^n$ are strictly positive. ■

With this observation in mind, we can restate the above basic result in a different way, as a purely as a theorem in linear algebra rather than a theorem in probability.

■ **Theorem:** Let $P$ be a stochastic matrix (finite order) such that for some $n$, all entries of $P^n$ are strictly positive. then $P^n$ converges to a stochastic matrix $\Pi$ which has identical rows; Further,

$$\Pi P = P\Pi = \Pi\Pi = \Pi. \qquad \blacksquare$$

Nice class, but few are irregular
and few come late, this is regular
Rules are not many, we have, for you
Trust is all what we have, in you.