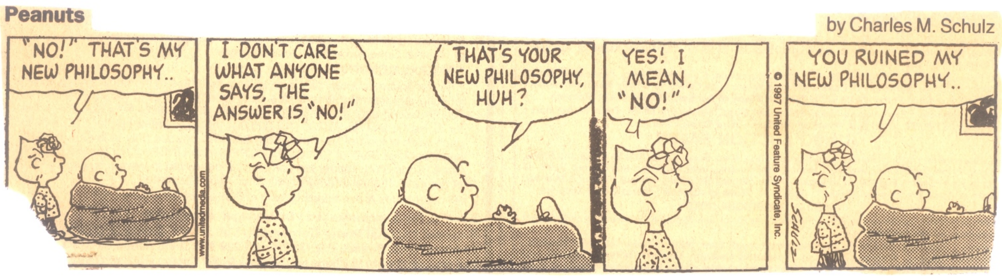

irreducible aperiodic:

Suppose we have a finite state MC with transition matrix P. Let i be a state. We defined

$$d_i = gcd \;\; A; \qquad A \;\; = \;\; \{n \geq 1 : p_{ii}^{(n)} > 0\}$$

Observe that

$$m, n \in A \quad \Rightarrow \quad m + n \in A \qquad (\bullet)$$

This is because

$$p_{ii}^{(m+n)} \geq p_{ii}^{(m)} p_{ii}^{(n)} > 0$$

Suppose that i is aperiodic, that is, the gcd above equals one. Then we claim:

$$\exists \;\; N \;\; \forall (n \geq N) \;\; n \in A.$$

Indeed we then have integers $n_1, \cdots, n_k$ and elements in A, say, $a_1, \cdots, a_k$ such that $\sum a_i n_i = 1$. Denoting

$$\sum_{n_i > 0} n_i a_i = x; \qquad \sum_{n_i < 0} a_i n_i = -y$$

we see, in view of $(\bullet)$ that $x, y \in A$. Given any integer n, we can express, $n = dy + r$ where $0 \leq r < y$. Let us now assume $n \geq y^2$, then we see $d \geq y$. Thus $d - r \geq 0$. Further,

$$n = dy + r = dy + rx - ry = rx + (d - r)y \in A$$

where we have used $(\bullet)$ again. Thus $N = y^2$ will do.

Hence, i is aperiodic implies there is an N such that $p_{ii}^{(n)} > 0$ for all $n \geq N$.

Let now $j \sim i$, that is, j communicates with i. We claim that j is also aperiodic. Indeed by definition of communication, fix $k, l \geq 1$ such that

$$p_{ij}^{(k)} > 0 \qquad\qquad p_{ji}^{(l)} > 0.$$

Let

$$e \quad = \quad gcd \;\; B; \quad B = \{n \geq 1 : p_{jj}^{(n)} > 0\}$$

If $n \in A$ then

$$p_{jj}^{(l+n+k)} \geq p_{ji}^{(l)} \; p_{ii}^{(n)} \; p_{ij}^{(k)} > 0$$

so that

$$l + n + k \in B; \qquad \text{or} \quad e \;\; \text{divides} \;\; l + n + k \quad (\star)$$

Also

$$p_{jj}^{(l+k)} \geq p_{ji}^{(l)} \;\; p_{ij}^{(k)} > 0$$

so that

$$l + k \in B; \qquad \text{or} \quad e \;\; \text{divides} \;\; l + k \quad (\star\star)$$

$$(\star)\;(\star\star) \quad \Rightarrow \quad e \quad \text{divides } n$$

Thus e divides every element of A and hence $e = 1$.

In fact the same argument shows the following:

$$i \quad \text{has period} \quad d \Rightarrow \exists N \;\; \forall (n \geq N) \;\; (nd \in A)$$

Actually, you need not repeat the proof, you can deduce this out of the previous argument as follows. Set

$$A = \{n \geq 1 : p_{ii}^{(n)} > 0\}; \qquad A^* = \{n/d : \quad n \in A\}$$

then A is closed under addition; hence A^* is also closed under addition; $gcd \;\; A^* = 1$; so after some stage all integers are in A^*; so after some stage all multiples of d are in A.

The same argument as in the aperiodic case above also shows the following. If period(i) $= d$ and $j \sim i$ then period(j) $=d$. If a property, that holds for one state of a communicating class, holds for all states in that class is called a **'class property'**.

Thus: *having Period d is class property.*

Returning to the aperiodic case, consider any aperiodic state i and $i \sim j$. We claim that there is an N such that $p_{ij}^{(n)} > 0$ for all $n \geq N$. Indeed there is N_1 such that $p_{ii}^{(n)} > 0$ for all $n \geq N_1$. Since $i \sim j$ fix k such that $p_{ij}^{(k)} > 0$. Clearly for all $n \geq N + k$ we have

$$p_{ij}^{(n+k)} \geq p_{ii}^{(n)} p_{ij}^{(k)} > 0$$

.

In particular we have the following:
If k is a aperiodic state in a finite state MC then every state in the communicating class $C(k)$ is aperiodic. Further there is on N such that for all $n \geq N$ and for all i, j in this class $p_{ij}^{(n)} > 0$.

This is because, there are only finitely many pairs (i, j) and the earlier para takes over.

In particular we have:
For a finite state irreducible aperiodic MC there is an N such that $p_{ij}^{(n)} > 0$ for all $n \geq N$ and for all states i, j.

As far as the matrices P^n are concerned, what this says is just the following:

If P is the transition matrix of an irreducible aperiodic MC, then after some stage onwards all entries of all P^n are strictly positive.

Of course if all the entries of *one* P^n are strictly positive then the chain is firstly irreducible. It is also aperiodic simply because then the next power of P also has strictly positive entries.

This concludes the proof of the statement made earlier, namely 'irreducible aperiodic' is equivalent to 'all entries of some power of P are strictly positive'.

Markov property/Elementary probability calculations:

Throughout S is a countable state space.
I shall omit taking care of a certain subtle point in order not to confuse you. This makes me appear little careless, let it be so.

let us consider a MC $\{X_n : n \geq 0\}$ with transition matrix P. Remember this simply means that there is some initial distribution $\mu = \{\mu_i : i \in S\}$ [sometimes written as $\mu(i)$] and

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

for all $n \geq 0$ and all states $i_0, \cdots , i_n$.

Actually, we have defined by two equations:

$$\begin{aligned} &P(X_0 = i_0) = \mu(i_0); \\ &P(X_n = i_n | X_0 = i_0; X_1 = i_1; \cdots ; X_{n-1} = i_{n-1}) = p_{i_{n-1}i_n} \end{aligned} \tag{2}$$

As you see these two descriptions are equivalent. In fact, if (1) holds then taking $n = 0$ you get first equation of (2) and definition of conditional probability gives you second equation of (2). Conversely if (2) holds then (1) holds simply by using multiplication rule

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1) \cdots P(A_n|A_1 \cap \cdots \cap A_{n-1})$$

We can, using (1), say

$$\begin{aligned} P(X_n = i_n | X_0 = i_0; X_1 = i_1; \cdots ; X_{n-1} = i_{n-1}) \\ = P(X_n = i_n | X_{n-1} = i_{n-1}) = p_{i_{n-1}i_n} \end{aligned} \tag{3}$$

In this form this property is known as **'Markov Property'**:

given all the past till today; state of tomorrow depends only on state of today.

Proof of (3) is easy. Let us make an observation first. Suppose A is an event and $\{B_i : i \leq k\}$ are disjoint events. suppose $P(A|B_i) = c$ for each i, then $P(A| \cup B_i) = c$. Proof is simple:

$$P(A \cap B_i) = cP(B_i). \quad P(A \cap [\cup B_i]) = P(\cup [A \cap B_i])$$

now by disjointness of $\{B_i\}$, we get

$$P(A \cap [\cup B_i]) = \sum P(A \cap B_i) = c \sum P(B_i) = cP(\cup B_i).$$

Returning to (3), Let us fix i_n and i_{n-1}. Take $A = \{X_n = i_n\}$; For each fixed $(n - 1)$ states $s = \langle i_0, \cdots , i_{n-2} \rangle$ consider the event

$$B_s = \{X_0 = i_0; \cdots ; X_{n-2} = i_{n-2}, X_{n-1} = i_{n-1}\}$$

What (1) says is that for each s

$$P(A|B_s) = c; \qquad c = p_{i_{n-1}i_n}$$

Thus

$$P(A|\bigcup_s B_s) = c; \qquad P(X_n = i_n|X_{n-1} = i_{n-1}) = c$$

which is what we wanted to prove because the events $\{B_s\}$ are disjoint and their union equals the event $\{X_{n-1} = i_{n-1}\}$.

It is necessary to give all the past? Given some precise past and state of yesterday, does state of tomorrow still depend only on state of yesterday? Yes. Instead of giving general formulae with large notation, we illustrate.

$$P(X_5 = i_5|X_1 = i_1, X_3 = i_3, X_4 = i_4) = P(X_5 = i_5|X_4 = i_4) = p_{i_4 i_5} \tag{4}$$

(Tomorrow is day 5; today is day 4; not all past is given)

equation (4) is easily proved. Bring in all the past info, that is express

$$(X_1 = i_1, X_3 = i_3, X_4 = i_4) = \bigcup_{s=\langle i_0 i_2\rangle} (X_j = i_j : 0 \leq j \leq 4)$$

and use the same argument as above.

Is it necessary to provide precise info of the remote past, any way future does not depend on the remote past? Suppose some imprecise past is given and precise state of yesterday is given. Does state of tomorrow still depend only on state of yesterday? Yes. Take sets of states E and F.

$$P(X_5 = i_5|X_1 \in E, X_3 \in F, X_4 = i_4) = P(X_5 = i_5|X_4 = i_4) = p_{i_4 i_5} \tag{5}$$

(You are given imprecise info about day one and three, not exact states)

You just express

$$(X_1 \in E, X_3 \in F, X_4 = i_4) = \bigcup_{i_1 \in E; i_3 \in F} (X_1 = i_1, X_3 = i_3, X_4 = i_4)$$

and use (4).

What if state of yesterday is not known? Some precise latest info is known, latest may be info of two days old? Yes, state of tomorrow depends on the latest state only.

$$P(X_7 = i_7|X_1 = i_1, X_3 = i_3, X_4 = i_4) = P(X_7 = i_7|X_4 = i_4) = p^{(3)}_{i_4 i_7} \tag{6}$$

Chances of being in i_7 on day 7 given the past as stated; equals chances of being in i_7 on third day starting at i_4. The past does not matter, you should reach i_7 in these three days starting at i_4.

By the way, in what follows, P^n is the n-th power of the matrix P and $p_{ij}^{(n)}$ is the (ij)-th entry of it.

To see (6), Let us observe a simple fact. For disjoint events (A_i)

$$P(\cup A_i | B) = \sum P(A_i | B)$$

For each i_5, i_6 states

$$P(X_7 = i_7; X_6 = i_6, X_5 = i_5 | X_1 = i_1, X_3 = i_3, X_4 = i_4) =$$

$$P(X_5 = i_5 | X_1 = i_1, X_3 = \imath_3, X_4 = i_4)$$
$$\times P(X_6 = i_6 | X_1 = i_1, X_3 = i_3, X_4 = i_4; X_5 = i_5)$$
$$\times P(X_7 = i_7 | X_1 = i_1, X_3 = i_3, X_4 = i_4; X_5 = i_5; X_6 = i_6)$$
$$= p_{i_4 i_5} p_{i_5 i_6} p_{i_6 i_7}$$

The first equality is multiplication rule; the last equality is from previous observation. summing both sides w.r.t. i_5, i_6 we get the stated result (6). You can think of variations of the above and prove easily. For example let E and F be two sets.

$$P(X_7 = i_7 | X_1 \in E, X_3 \in F, X_4 = i_4) = P(X_7 = i_7 | X_4 = i_4) = p_{i_4 i_7}^{(3)} \tag{7}$$

The upshot is the following: Tomorrow's state depends only on the latest state. Thus if you know some info about states on days 5,8,12 and precise state of day 15 (say it is i) then chances of being in state j on day 21 is simply being there at j in six days starting from i, that is, $p_{ij}^{(6)}$.

What about not just tomorrow but next two days?

$$\begin{aligned} P(X_8 = i_8; X_7 = i_7 | X_1 \in E, X_3 \in F, X_4 = i_4) \\ = P(X_7 = i_7 | X_4 = i_4) = p_{i_4 i_7}^{(3)} p_{i_7 i_8} \end{aligned} \tag{8}$$

This follows from multiplication rule and the previous result.

In fact you can generalize this in several ways. We state one more and leave the rest to your imagination.

$$\begin{aligned} P(X_{20} = i_{20}; X_{15} = i_{15}; X_9 = i_9 | X_1 \in E, X_3 \in F, X_4 = i_4) \\ = P(X_{20} = i_{20}; X_{15} = i_{15}; X_9 = i_9 | X_4 = i_4) \\ = p^{(5)}_{i_4 i_9} p^{(6)}_{i_9 i_{15}} p^{(5)}_{i_{15} i_{20}} \end{aligned} \tag{9}$$

You should think of this as follows. Chances of the event

$$\{X_{20} = i_{20}; X_{15} = i_{15}; X_9 = i_9\}$$

given the event

$$(X_1 \in E, X_3 \in F, X_4 = i_4)$$

is same as the following.

chance of: starting at i_4 in five days go to i_9; from then in 6 days go to i_{15} and from there in 5 days go to i_{20}.

on communicating classes:

Communicating classes do not depend on the initial distribution. They depend only on the transition matrix. In other words they are determined by the ‘law of motion‘ and do not depend on how you started the motion.

We saw that ‘period d’ is a class property; if one state is so then all other states in its class are also so. there are other properties which we shall seed later.

review:

There seems to be some difficulty in understanding MC. This is the first time we are dealing with infinite sequence of random variables ‘at a time’ and questions are of the type ‘what happens ultimately’. You should appreciate that we are able to answer complicated questions which are of great importance. We discussed examples before starting on the theory.

We started with Chandrasekhar model: balls are added and removed to the urn. This arose in connection with diffusion and Brownian motion as explained there. we were able to calculate what happens ultimately, lucky.

Then we discussed (simple symmetric) random walk. You are sure to return to origin, starting point, in one/two dimensions and in three dimensions there is a non-zero chance of not returning. This, apart from other applications, is related to brownian motion. What happens if you move more frequently and make small moves? You get, if properly formulated, Brownian Motion.

Then we discussed the two state MC. There were several scenarios.

Some times there was an absorbing state;

sometimes there is a stationary distribution which is equilibrium distribution, that is, no matter how you start your ultimate fate is this distribution.

sometimes there is a stationary distribution but it is not equilibrium distribution. Thus when you started in some state you did not end up with this stationary distribution.

There are some general principles involved in these calculations and the theory of Markov Chains isolates these and builds a theory.

basic theorem:

For a finite state irreducible aperiodic MC we stated a basic theorem with several parts. Actually existence of limit of P^n is the essence. From that other parts follow.

For example, the fact that the limit Π is stochastic is immediate. Row of each P^n adds to one, the same holds in limit because it is a finite sum (important: finite state space!).

Take a row π of the matrix Π. then we have $\pi P = \pi$. in fact

$$\Pi P = (\lim P^n)P = \lim(P^n P) = \lim P^{n+1} = \Pi$$

Thus each row of Π is a stationary distribution.

We shall now show that the solution space:

$$\{v \in R^{|S|} : vP = v\} \qquad (\spadesuit)$$

is one dimensional. then it follows that there is at most one probability vector π with $\pi P = \pi$. Since each row of Π satisfies this equation, we conclude that the matrix Π has identical rows.

Using the fact that row rank equals column rank, to show (♠) it is enough to show that the space

$$\{w \in R^{|S|} : Pw = w\} \qquad (♠)$$

is one dimensional. (Place value tells you if it is row vector or column vector). Since constant vector $w \equiv 1$ is a solution; we need to show that every solution is a multiple of this.

Let w be a solution and let M be the maximum of its coordinates and let k be a coordinate such that $w_k = M$. Now, $w = Pw$ tells us

$$M = w_k = \sum_i p_{ki} w_i \leq M$$

which, in particular says

$$p_{ki} > 0 \quad \Rightarrow \quad w_i = M.$$

Since $w = Pw$ implies that $w = P^n w$ for each n, the same argument shows the following. suppose i is a state with $p_{ki}^{(n)} > 0$ then $w_i = M$. But the chain being irreducible such an n exists for every i. Hence $w_i = M$ for every i.

Observe that we did not need to use aperiodicity in the proof above. Thus we have the following.

If the stochastic matrix P is irreducible then every solution of $Pw = w$ is a constant vector. There is only one probability vector π such that $\pi P = \pi$.

Do not confuse, we never said that left eigen vector is a constant vector, in general not true. We only said right riven vector is constant vector.

returning to our basic theorem, since every row of the limit matrix is a stationary distribution, the above uniqueness tells you Π has identical rows.

This argument shows that all parts of the basic theorem are consequences of its first sentence, namely existence of limit of P^n.

Card shuffling:

here is the exercise about card shuffling: Pick a card at random from the stack and put it on the top. here each day, state of the system is the arrangement of the cards. Thus state space is the set of permutations of the 52 cards. hence $|S| = 52!$. Thus the transition matrix is a $(52! \times 52!)$ matrix.

Form one state we can go to 52 other states depending on which card is picked. Thus the matrix has only 52 non-zero entries in each row. Further each non-zero entry is same: $1/52$.

If you pick first card then the you do not change the state. This should convince you that each state is aperiodic.

is it irreducible? From any state, by a sequence of applications of this process you can go to the identity permutation and from identity permutation you can reach any other permutation. (Do you see?) Hence the chain is irreducible.

So the fundamental theorem tells you that our huge matrix has its powers converging to some matrix. What is the limit? Since the limit matrix consists of identical rows, you need to see one row. Since each row is a stationary distribution you need to see which probability vectors satisfy the equation:

$$\mu P = \mu$$

(Unfortunately π is the standard notation for permutations, so I can not use for stationary distribution now. We have enough greek letters!)

Can we solve this equation? Trivial if only you think! observe that our matrix also has the following extra property: Every column also has 52 non-zero entries and each non-zero entry is $1/52$. Given a permutation, what are the permutations that lead to this? Think about it (and make sure you are capable of writing the argument).

In other words our matrix has non-negative entries and each row sum as well as each column sum equal one. Such matrices are called **doubly stochastic matrices**.

For such a matrix constant vector is also a left eigen vector. Thus one solution for $\mu P = \mu$ is

$$\mu(s) = \frac{1}{52!}; \qquad\qquad s \in S.$$

Thus limiting matrix consists of this as each of its rows. But of course, before saying this, you need to make sure that this is the only solution. Yes, uniqueness of stationary distribution tells us this.

In other words the uniform distribution is the stationary distribution. Basic theorem tells us if you run the chain for a long time and ask the

question: what are the chances now I am in state s ?

the answer is: nearly 1/52!

In other words, each state is equally likely; you achieved uniform arrangement of the cards. This is precisely the purpose of shuffling cards. This is a very slow process. There are other ways of shuffling like riffle shuffle. We shall not visit this.

In the process of this argument, we have discovered that for a doubly stochastic matrix uniform distribution is a stationary distribution; if it is irreducible then it is is the only stationary distribution.

Knight moves:

The problem on moves of the knight on the chess board is interesting. It brings in ideas that are applicable more generally.

Here the state space is the 64 squares of the chess board. Thus $|S| = 64$. You can name the states any way you like. But if you name them as $\{(x, y) : 0 \leq x, y \leq 7\}$ then you see that any one move changes the parity of $x + y$ and hence every state has period 2.

You can reach bottom row from any square; Further from any square in bottom row you can reach its first square.

[proceed only if you have verified the statement].

Since the moves are reversible, we conclude the following.

we can reach first square from any square

and reach any square from first square.

This is enough to conclude that the chain is irreducible (why?).

Thus the chain is irreducible but not aperiodic. Basic theorem is not applicable. Remember, that if there is a stationary distribution then it is unique. We have noted that proof of this fact does not need aperiodicity.

So the question is:

is there a stationary distribution? Since probably it is not obtained via $\lim P^n$, what does it convey to us?

Basic theorem part 2:

Let P be transition matrix of a finite state markov chain (X_n). Assume the chain is irreducible.

(1) Then there is a matrix Π such that

$$\frac{I + P + P^2 + \cdots + P^{n-1}}{n} \to \Pi$$

(as usual, entry wise).

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

(3) π is a probability vector.

(4) π satisfies $\pi P = \pi$.

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

(6) $\pi(s)$ is the average time spent by the chain in state s.

(7) $\pi(s) = 1/m_s$ where m_s is the expected time to return to s starting from s.

∎

In other words all the statements remain correct without aperiodicity except the first one; instead of the sequence (P^n), their averages have a limit. Further π has nice interpretation.

In any one run of the chain, we can ask the proportion of time spent in a state, say i, during the first n days. If this has limit then we can say that this is the proportion of time spent in that state i 'during the entire run'. More over if this limit does not depend on the run; then we can say that the chain spends this proportion in the state i without referring to the run. In our case this happens. For 'almost all' runs this proportion exists and equals $\pi(i)$.

Also just like in random walk, it makes sense to ask: if I started from i do I surely return to i? The answer is 'yes' for the chain under consideration (irreducible). Then it makes sense to ask what is the distribution of the time taken to return¿ It makes sense to ask if this return time has finite expectation. In our case the answer is the following: starting from i the time needed to return has expected value $1/\pi(i)$.

Graph walk:

Both the examples above, card shuffling and Knight moves have the following feature: Associated with each state there is a set of accessible states. When you are in a state you select one of its accessible states at random and move there.

In the card shuffling, for one possible arrangement of cards, there is a set of 52 possible arrangements which are accessible. In the knight moves, from one square, there are some squares (2 or 3 or 5 or 8) that are accessible.

This idea can be generalized. Consider a graph (V, E) allow loops, but no multiple edges. Here is a walk on the graph. State space is the set of vertices. from a vertex, you move to its neighbour at random. More precisely, for each vertex v, let $A(v)$ denote all vertices connected to v. If there is a loop at v then v will be in this set, otherwise no. Denote $d(v) = |A(v)|$ cardinality of the set $A(v)$. This is called degree of the vertex v.

if you are at v, choose a vertex at random from $A(v)$ and move there. Thus each vertex in $A(v)$ has chance $1/d(v)$ of being selected.

Let us assume that the graph is connected. then the chain is irreducible. For this chain the vector $\{d(v) : v \in V\}$ solves $\pi P = \pi$. Indeed, let us fix a vertex w

$$\sum_v d(v)P(v,w) = \sum_{v:v\in A(w)} d(v)P(v,w)$$

Note that $v \notin A(w)$ then $w \notin A(v)$ and hence $P(v,w) = 0$ so that the sum is over only those v showed above. But then for any such v, we have $P(v,w) = 1/d(v)$. Thus

$$\sum_v d(v)P(v,w) = \sum_{v:v\in A(w)} d(v)\frac{1}{d(v)}$$

$$= \sum_{v:v\in A(w)} 1 \quad = d(w)$$

Thus if $z = \sum\limits_v d(v) =$ twice number of edges in the graph, then

$$\pi = \left\{ \frac{d(v)}{z} : z \in V \right\}$$

is a probability vector and $\pi P = \pi$. So this is the stationary probability vector.

It is worthwhile recalling why it is called stationary. If you run the markov chain with this initial distribution, then every day the distribution of the state remains same as this π; does not change; remains stationary.

Returning to the chess problem, the stationary distribution can easily be calculated. do it.

Ehrenfests:

Rather than discussing generalities about Markov chains, I shall discuss one more model.

I have a box which has two compartments: I and II. It has $2R$ balls numbered $1, 2, \cdots, 2R$; some in I and some in II. here is the law of motion: Pick one of the $2R$ numbers at random and move that ball from its compartment to the other compartment.

Thus if you selected 5, see for ball number 5; if it is in compartment I take it out and put in II; if that ball is in II then put it in I.

Do it again and again. What happens in the long run?

Let us denote by X_n the number of balls in I after n exchanges; thus X_0 is the initial number.

The state space here is $\{0, 1, 2, \cdots, 2R\}$.

The transition matrix is given as follows.

$$p_{i,i+1} = \frac{2R - i}{2R}; \quad p_{i,i-1} = \frac{i}{2R}$$

$$p_{i,j} = 0; \quad \text{for} \quad j \neq i \pm 1.$$

This is because of the following reason. If there are i balls, then it increases only when you pick a number from II and put that ball in I. you know that now II has $(2R - i)$ balls out of the $2R$.

The number reduces when the number selected is that of a ball in I.

Since we did not want to separate out cases we have given the above formulae. You should remember that, for example when $i = 0$ then $p_{i,i-1} = p_{0,-1}$ does not make sense. But do not worry, our formula gives zero value. When $i = 0$ then surely you move to 1. Similarly from $2R$ you surely move to $2R-1$.

Just like random walk, the period is two. The chain is irreducible: from any state you can go to zero and from zero you can go to any state.

Due to periodicity P^n does not converge. What is the stationary distribution?

$$\pi(i) = \binom{2R}{i} \frac{1}{2^{2R}}; \qquad\qquad i = 0, 1, 2, \cdots, 2R.$$

In other words the binomial probabilities $B(2R, 1/2)$. In other words if you distribute at random each ball initially, then on each day it appears as if the balls are distributed at random! To show that it is stationary is easy. Fix j.

$$2^{2R} \sum \pi(i) P_{ij} = \pi(j-1) P_{j-1,j} \ + \ \pi(j+1) P_{j+1,j}$$

$$= \frac{(2R)|}{(j-1)!(2R-j+1)!} \ \frac{2R-j+1}{2R} \ + \ \frac{(2R)!}{(j+1)!(2R-j-1)!} \ \frac{j+1}{2R}$$

$$\binom{2R}{j} \left[\frac{j}{2R} \ + \ \frac{2R-j}{2R} \right]$$

$$= \pi(j) 2^{2R}.$$

Of course, $1/\pi(i)$ has the interpretation: starting from i the expected time to return to i. For example starting from 1 it takes

$$2^{2R} \frac{1}{2R}$$

time units, a huge number if R is huge.

We could have explained this with R balls instead of $2R$, there is no significance of 2. Historically, it was taken so.

What is going on and what is this model for?

Well, this is a model proposed by the husband-wife team of physicists Paul Ehrenfest and Tatiana Ehrenfest. If you mix hot milk and cold water what happens? the water seems to gain heat from milk and finally a steady temperature is reached. How is this happening? How is this giving-taking of heat going on?

Firstly, matter is made of small particles, call them atoms or whatever (the future nobel laureate Ernst Mach did not agree at that time, can you believe?) Heat is simply manifestation of the motion. The fast they move, more is the heat. When two materials of different temp mix, the particles collide with each other, exchange the speed/momentum/energy (call it anything that you are happy with; after all this is not physics course). After several collisions a uniform speed is reached.

But this exchange of energy is not a one sided one, not an orderly process, it is a complex process. A hot milk molecule may collide with cold water molecule and impart some velocity; later this fast moving water molecule may collide with a slow moving milk molecule and impart some energy. Thus huge number of collisions and energy exchanges take place in the process.

so what do the balls indicate. well the compartments are milk and water. *Number of balls signify temperature of that compartment.* More balls higher temperature. Less balls low temperature.

The exchange of ball signifies a collision. If you placed ball in I, it means a particle of I collided with a particle of II and gained energy thus increasing temperature (balls) in I.

Ultimately after all these exchanges what happens?

A steady state is reached. (♠)

It is $B(2R, 1/2)$ What is the expected number of balls in I after a long time? Mean of $B(n, p)$ variable is np. Thus ultimately mean number of particles in I equals

$$2R \times \frac{1}{2} = R$$

thus each compartment would have R balls, on the average. Same temperature? [Of course due to periodicity what I said is not exactly true, if you randomise initial state it is true.]

Let us consider ONE run of the game, one entire (for ever) performance of the motion. That is let us consider one sample point ω and see the values of

$$X_1(\omega), \quad X_2(\omega), \quad X_3(\omega), \quad \cdots$$

We can show that each of the numbers $0, 1, 2 \cdots, 2R$ appear infinitely many times in the above sequence.

In any particular run, each state appears infinitely many times. (♣)

(This is not entirely true. Strictly speaking, above phenomenon may not hold for all runs but it holds surely; that is with probability one; that is those runs for which the above fails has probability zero. You need not attend to such finer details now.)

This seems to say that each run is a chaotic thing, no particular limit or steadiness in the observations simply because no matter how far you go you see each of these states.

Thus (♠) and (♣) appear to be contradictory. (♠) says steady state is being reached, there is a limiting state for the system; (♣) says each run is chaotic, there is no limiting state. This lead to much confusion in the initial stages when Boltzmann proposed his theory. It is precisely to make people understand that there is no contradiction, Ehrenfests proposed this model.

Two things are to be noted.

Firstly, (♠) refers to what is happening at 'macro level'; it does not refer to specific state of the day; it refers to the distribution of the day and its limit. (♣) refers to what is happening at 'micro level'; it says about state of each day for each specific run of the system; for each possible motion of the system.

Secondly, the time to return to an unlikely state is so huge, you do not see in your life time!

To bring home the point of (♣) and appreciate why people did not believe Boltzmann, consider the following. Suppose you started with $X_0 = 1$; thus one ball in I and $(2R - 1)$ balls in II; thus cold water in I and hot milk in II. In any single run of this game of collisions, (♣) says, the state 1 appears again and again. That means, explanation of heat exchange by this method tells you after a long long time, after many many collisions, you are sure to see water as cold as at the beginning and milk as hot as at the beginning. This appeared unbelievable. What people did not realize is that the number R is so Huge, that the recurrence time is too Huge to be observable in one's life time.

Think about matters carefully.