The first step analysis

1 Some important definitions.

Let $X_n,\ n = 0, 1, 2, ...$ be a sequence random variables takin values in $S = (1, 2, ...m)$. Such sequences are often called random processes (RP). S is called the state space.

Definition 1.1
A RP $X_n,\ n = 0, 1, 2, ...$ is called a Markov chain (MC) if

$$P\{X_{n+1} = j \,|\, X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\} = P\{X_{n+1} = j \,|\, X_n = i\} \qquad (1)$$

Relation (**??**) is called *the Markov property*. It implies that once X_n is given the future development of the MC does not depend on what has happened before time n. This is sometimes stated as "the future and the past are independent if the present is known".
The probabilities $P\{X_{n+1} = j \,|\, X_n = i\}$ are called *the transition probabilities*. If these probabilities do not depend on n then the MC X_n is said to be *homogeneous* (or more precisely, *homogeneous in time*). In this course we shall study only homogeneous MCs. A homogeneous MC is basically described by its transition probabilities $p_{ij} \stackrel{\text{def}}{=} P\{X_{n+1} = j \,|\, X_n = i\}$ and it is convenient and useful to present the collection of these probabilities as a *transition matrix* of the MC X_n:

$$\mathbb{P} = (p_{ij}) = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{pmatrix}$$

2 A reminder

We recall the following general formula for an expectation of a random variable known from Probability II. Let ξ be a random variable and let events $B_1, B_2, ..., B_m$ be such that $P(B_1) + ... + P(B_m) = 1$ and $B_j \bigcap B_k = \emptyset$ if $j \neq k$. Then

$$E(\xi) = \sum_{j=1}^{m} P(B_j) E(\xi \,|\, B_j). \qquad (2)$$

In particular, if η is a random variable defined on a set of trajectories of a MC then

$$E(\eta \,|\, X_0 = i) = \sum_{j=1}^{m} p_{ij} E(\eta \,|\, X_0 = i, X_1 = j). \qquad (3)$$

Note that (**??**) is obtained from (**??**) with $B_j = \{X_1 = j, \,|\, X_0 = i\}$ and $\xi = (\eta \,|\, X_0 = i)$.

3 First step analysis (FSA)

We say that a state i of a MC X_n is *absorbing* if $p_{ii} = 1$.
Suppose that a MC X_n has absorbing states and let T be the time at which it reaches (is absorbed by) one of these states. If the MC eventually reaches one of the absorbing states (or, equivalently, if T is finite) then we say that X_n is an *absorbing MC*.

Remark. Strictly speaking, X_n is an *absorbing MC* if T is finite with probability 1. However, this level of rigor is beyond the technical means of our course.
Let $f(i)$ be a function on S taking real values and set

$$F = f(X_0) + f(X_1) + ... + f(X_T) \equiv \sum_{j=0}^{T} f(X_j).$$

FSA is a technique which allows one to find $E(F \,|\, X_0 = i)$ by establishing a relation between this quantity and the one into which this expectation is transformed after the MC makes its first step, namely the $E(F \,|\, X_0 = i, X_1 = j)$. Set

$$w_i = E(f(X_0) + f(X_1) + ... + f(X_T) \,|\, X_0 = i) \equiv E(F \,|\, X_0 = i).$$

The FSA allows one to prove the following

Theorem 3.1
Suppose that X_n is an absorbing MC. Then

$$\begin{cases} w_i = & f(i) \quad \text{if } i \text{ is absorbing} \\ w_i = & f(i) + \sum_{j=1}^{m} p_{ij} w_j \quad \text{if } i \text{ is not absorbing} \end{cases} \tag{4}$$

and these equation have a unique solution.

We shall prove that equations (**??**) hold. The existence and uniqueness of their solutions will not be proved.
Proof. The proof of the first equation is immediate since if i is absorbing then $(F \,|\, X_0 = i) = f(i)$ and hence $w_i = E(F \,|\, X_0 = i) = f(i)$.
To prove the second equation we start with formula (**??**) with $\eta = F$. Observe that then $E(\eta \,|\, X_0 = i, X_1 = j)$ becomes

$$\begin{aligned} E(F \,|\, X_0 = i, X_1 = j) &= f(i) + E(f(X_1) + f(X_2) + ... + f(X_T) \,|\, X_1 = j) \\ &= f(i) + E(f(X_0) + f(X_1) + ... + f(X_T) \,|\, X_0 = j). \end{aligned}$$

The last step in this formula is due to the fact that we deal with a homogeneous MC and therefore the distribution of $(f(X_1) + f(X_2) + ... + f(X_T) \,|\, X_1 = j)$ is the same as that of $(f(X_0) + f(X_1) + ... + f(X_T) \,|\, X_0 = j)$. Hence

$$E(f(X_1)+f(X_2)+...+f(X_T) \,|\, X_1 = j) = E(f(X_0)+f(X_1)+...+f(X_T) \,|\, X_0 = j) \equiv E(F \,|\, X_0 = j)$$

and thus

$$E(F \,|\, X_0 = i, X_1 = j) = f(i) + w_j.$$

Substituting this expression into the r.h.s. of (**??**) and replacing the l.h.s. of (**??**) by w_i we obtain

$$w_i = \sum_{j=1}^{m} p_{ij}(f(i) + w_j) = f(i)\sum_{j=1}^{m} p_{ij} + \sum_{j=1}^{m} p_{ij}w_j. \tag{5}$$

Since $\sum_{j=1}^{m} p_{ij} = 1$, we see that (**??**) implies (**??**) and this finishes the proof of the theorem. □

4 Examples.

The examples presented below differ by the choice of the function f.

1) Set $f(i) = \begin{cases} 0 & \text{if } i \text{ is an absorbing state} \\ 1 & \text{if } i \text{ is not an absorbing stae} \end{cases}$

Then $F = f(X_0) + f(X_1) + ... + f(X_T) = f(X_0) + f(X_1) + ... + f(X_{T-1}) = T$. Indeed, X_T is an absorbing state by the definition of T whereas X_r, $0 \le r \le T-1$, is not. Hence $f(X_T) = 0$ and $f(X_0) = f(X_1) = ... = f(X_{T-1} = 1$. We recover the equations for $v_i \stackrel{\text{def}}{=} E\{T \,|\, X_0 = i\}$ (derived directly from the definition of T in one of previous lectures), namely

$$\begin{cases} v_i = & 0 \quad \text{if } i \text{ is absorbing} \\ i_i = & 1 + \sum_{j=1}^{m} p_{ij}v_j \quad \text{if } i \text{ is not absorbing} \end{cases} \tag{6}$$

2) For a fixed non-absorbing state k set $f(i) = \begin{cases} 1 & \text{if } i = k \\ 0 & \text{in all other cases} \end{cases}$ Then

$F = f(X_0)+f(X_1)+...+f(X_T) = f(X_0)+f(X_1)+...+f(X_{T-1}) =$ the number of visits to k.

Indeed, $f(X_T) = 0$ since X_T is an absorbing state and for $f(X_r) = 1$ if and only if $X_r = k$. Hence the equality.

Let us denote by a the number of visits to k and set $b_i \stackrel{\text{def}}{=} E\{a \,|\, X_0 = i\}$. Then b_i can be found from the following equations

$$\begin{cases} b_i = & 0 \quad \text{if } i \text{ is absorbing} \\ b_i = & \delta_{ik} + \sum_{j=1}^{m} p_{ij}b_j \quad \text{if } i \text{ is not absorbing} \end{cases} \tag{7}$$

(Remember that $\delta_{ik} \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } i = k \\ 0 & \text{if } i \ne k \end{cases}$)

3) For a fixed absorbing state k set $f(i) = \begin{cases} 1 & \text{if } i = k \\ 0 & \text{in all other cases} \end{cases}$ Then

$$F = f(X_0) + f(X_1) + ... + f(X_T) = f(X_T) = \begin{cases} 1 & \text{if } X_T = k \\ 0 & \text{if } X_T \ne k. \end{cases}$$

Indeed, $f(X_0) = f(X_1) = ... = f(X_{T-1} = 0$ because $X_0, ..., X_{T-1}$ are not absorbing states and $f(X_T) = 1$ if and only if $X_T = k$.
Note next that in this case $u_i = E(F \mid X_0 = i) = E\{X_T \mid X_0 = i\} = P\{X_T = k, \mid X_0 = i\}$. Once again, we recover the equations for u_i (derived directly from the definition of u_i in one of previous lectures), namely

$$\begin{cases} u_i = & \delta_{ik} \text{ if } i \text{ is absorbing} \\ u_i = & \sum_{j=1}^{m} p_{ij} u_j \text{ if } i \text{ is not absorbing} \end{cases} \tag{8}$$

4) If a game is described by our MC, $f(i)$ is sometimes interpreted as a reward for visiting state i (which can perfectly well be a negative number) and thus $F = \sum_{k=0}^{T} f(X_k)$ is the reward paid out at the end of the game. The w_i is then the expectation (the mean value) of the reward given that the starting point of the process is i. If the initial distribution $\pi = (p_1, ..., p_m)$ of the chain is known then we can also find the $E(F) = \sum_{i=1}^{m} p_i w_i$.
In some cases it is natural to say that a game is fair if $E(F) = 0$. Solving equations (**??**) enables one to find out whether or not a particular game is fare.

We shall consider much more concrete examples in lectures.