In fact, it can be shown that the $p_j,\ j \in \mathcal{G}$ are the unique nonnegative solutions of equations (121) and (122). All this is summed up in Theorem 30, which we state without proof.

Theorem 30. For an ergodic Markov chain $p_j = \lim_{n\to\infty} P_{i,j}(n)$ exists, and the $p_j,\ j \in \mathcal{G}$ are the unique nonnegative solutions of equations (121) and (122).

Returning to the Markov chain of Example 97, we see that Theorem 30 does not apply. Although the limits

$$\lim_{n\to\infty} p_{i,j}(n)$$

of $p_{i,j}(n)$ exist (in fact $\P(n) = \P$) and is not dependent on the initial distribution.

We consider the two–state Markov chain. From this we conclude that

$$p_1 = \frac{p_{2,1}}{p_{1,2} + p_{2,1}} \quad \text{and} \quad p_2 = \frac{p_{1,2}}{p_{1,2} + p_{2,1}}. \tag{123}$$

This result can also be derived using the system (121), (122). Using (121) we obtain

$$p_1 = (1 - p_{1,2})\, p_1 + p_{2,1}\, p_2$$

$$p_2 = p_{1,2}\, p_1 + (1 - p_{2,1})\, p_2$$

Note that these two equations are linearly dependent, and thus we need one more equation (supplied by the condition (122)):

$$p_1 + p_2 = 1.$$

Solving we get the same limiting probability distribution (123).

Example 98. Let for a Markov chain with three states and the one-step transition probabilities matrix:

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

It is not difficult to verify that this Markov chain is ergodic, since elements of the two-step transition probabilities matrix are all positive (condition (120)). Using (121) we obtain

$$2\,p_1 = p_2 + p_3,$$

$$2\,p_2 = p_1 + p_3,$$

and

$$2\,p_3 = p_2 + p_3.$$

Solving the system we obtain

$$p_1 = p_2 = p_3$$

Using (122) we get $p_1 = p_2 = p_3 = 1/3$.

Example 99. Consider Markov chain:

$$¶ = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{2}{3} & 0 & \frac{1}{3} \\ \frac{3}{5} & \frac{2}{5} & 0 \end{pmatrix}$$

This Markov chain is ergodic, since all elements of the two-step transition probabilities matrix are positive (see Example 89). Using (121) and (122) we obtain the following system of linear equations:

$$p_1 = 1/2\,p_1 + 2/3\,p_2 + 3/5\,p_3,$$

$$p_2 = 1/4\,p_1 + 2/5\,p_3,$$

$$p_3 = 1/4\,p_1 + 1/3\,p_2,$$

and

$$p_1 + p_2 + p_3 = 1.$$

Solving the system we obtain

$$p_1 = \frac{52}{93} \qquad p_2 = \frac{21}{93} \qquad p_3 = \frac{20}{93}.$$

Theorem 27 (Chapman–Kolmogorov equations).

$$P_{i,j}(n) = \sum_{k \in \mathcal{G}} P_{i,k}(r) \cdot P_{k,j}(n-r) \tag{116}$$

for all $0 < r < n$.

The Chapman–Kolmogorov equations (116) provide a method for computing n–step transition probabilities. Equations (116) are most easily understood by noting that $P_{i,k}(r) \cdot P_{k,j}(n-r)$ represents the probability that starting in i the process will go to state j in n transitions through a path which takes it into state k at the rth transition. Hence, summing over all intermediate states k yields the probability that the process will be in state j after n transitions. Formally, we have

$$P_{i,j}(n) = P(\eta_n(\omega) = j \,/ \eta_0(\omega) = i\,) = \sum_{k \in \mathcal{G}} P(\eta_n(\omega) = j, \eta_r(\omega) = k \,/ \eta_0(\omega) = i\,) =$$
$$= \sum_{k \in \mathcal{G}} P(\eta_r(\omega) = k \,/ \eta_0(\omega) = i\,) \cdot P(\eta_n(\omega) = j \,/ \eta_r(\omega) = k\,) = \sum_{k \in \mathcal{G}} P_{i,k}(r) \cdot P_{k,j}(n-r).$$

If we let $\P(n)$ denote the matrix of n–step transition probabilities $P_{i,j}(n)$, then equation (116) asserts that

$$\P(n) = \P(r) \cdot \P(n-r),$$

where the dot represents matrix multiplication[1]. Hence, in particular $(n = 2,\ r = 1)$

$$\P(2) = \P \cdot \P = \P^2$$

and by induction

$$\P(n) = \P^{n-1} \cdot \P = \P^n.$$

That is, the n–step transition probability matrix may be obtained by multiplying the matrix $\P$ by itself n times.

[1]If $\mathbf{A}$ is an $n \times m$ matrix whose element in the ith row and jth column is a_{ij} and $\mathbf{B}$ is an $m \times k$ matrix whose element in the ith row and jth column is b_{ij}, then $\mathbf{A} \cdot \mathbf{B}$ is defined to be the $n \times k$ matrix whose element in the ith row and jth column is $\sum\limits_{k=1}^{m} a_{ik} \cdot b_{kj}$

Example 83. Let $\{\eta_n(\omega); n \in \mathbb{N}\}$ be a Markov chain with three states and transition matrix

$$\P = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{2}{3} & 0 & \frac{1}{3} \\ \frac{3}{5} & \frac{2}{5} & 0 \end{pmatrix}$$

Then

$$P(\eta_1 = 2 \cap \eta_2 = 3 \cap \eta_3 = 1 \cap \eta_4 = 3 \cap \eta_5 = 1 \cap \eta_6 = 3 \cap \eta_7 = 2 / \eta_0 = 3) =$$

$$= p_{3,2} \cdot p_{2,3} \cdot p_{3,1} \cdot p_{1,3} \cdot p_{3,1} \cdot p_{1,3} \cdot p_{3,2} = \frac{2}{5} \cdot \frac{1}{3} \cdot \frac{3}{5} \cdot \frac{1}{4} \cdot \frac{3}{5} \cdot \frac{1}{4} \cdot \frac{2}{5} = \frac{3}{2,500}.$$

The two–step transition probabilities are given by

$$\P(2) = \P^2 = \begin{pmatrix} \frac{17}{30} & \frac{9}{40} & \frac{5}{24} \\ \frac{8}{15} & \frac{3}{10} & \frac{1}{6} \\ \frac{17}{30} & \frac{3}{20} & \frac{17}{60} \end{pmatrix}$$

so that, for example $P_{2,3}(2) = \frac{1}{6}$.

Example 84 (Number of successes in Bernoulli process). Let $\eta_n(\omega)$ denote the number of successes in n independent trials, where the probability of a success in any one trial is p. The sequence $\{\eta_n(\omega); n \in \mathbb{N}\}$ is a Markov chain. Here the state space is $\{0, 1, 2, ...\}$, the initial distribution is $\pi(0) = 1$ and $\pi(j) = 0$ for $j \geq 1$. The transition matrix is

$$\begin{pmatrix} 1-p & p & 0 & ... & 0 \\ 0 & 1-p & p & ... & 0 \\ 0 & 0 & 1-p & ... & 0 \\ 0 & ... & 0 & 1-p & p \end{pmatrix}$$

If the state space $\mathcal{G}$ of a Markov chain is finite, then computing $P_{i,j}(n)$ is relatively straightforward.

Example 85. We observe the state of a system at discrete points in time. We say that the system is in state 1 if it is operating properly. If the system is undergoing repair (following a breakdown), then the system state is denoted by state 2. If we assume that the system possesses the Markov property, then we have a two-state (homogeneous) Markov chain. We have

$$\P = \begin{pmatrix} 1 - p_{1,2} & p_{1,2} \\ p_{2,1} & 1 - p_{2,1} \end{pmatrix}$$

In this particular case we can compute $P_{i,j}(n)$. We will impose the condition

$$|1 - p_{1,2} - p_{2,1}| \neq 1 \quad \text{or} \quad < 1 \tag{117}$$

on the one–step transition probabilities.

If $|1 - p_{1,2} - p_{2,1}| < 1$, then n–step transition probability matrix $\P(n) = \P^n$ is given by:

$$\P(n) = \begin{pmatrix} \frac{p_{2,1} + p_{1,2}\,(1 - p_{1,2} - p_{2,1})^n}{p_{1,2} + p_{2,1}} & \frac{p_{1,2} - p_{1,2}\,(1 - p_{1,2} - p_{2,1})^n}{p_{1,2} + p_{2,1}} \\ \frac{p_{2,1} - p_{2,1}\,(1 - p_{1,2} - p_{2,1})^n}{p_{1,2} + p_{2,1}} & \frac{p_{1,2} + p_{2,1}\,(1 - p_{1,2} - p_{2,1})^n}{p_{1,2} + p_{2,1}} \end{pmatrix} \tag{118}$$

Since $p_{1,2}$ and $p_{2,1}$ are probabilities, condition (117) can be violated only if $p_{1,2} = p_{2,1} = 0$ or $p_{1,2} = p_{2,1} = 1$. These two cases is treated separately.

Example 86. Let $p_{1,2} = p_{2,1} = 0$. Clearly, $|1 - p_{1,2} - p_{2,1}| = 1$, and therefore (118) does not apply. The transition probability matrix $\P$ is the identity matrix:

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

The two states do not communicate with each other. The matrix $\P(n) = \P^n$ is easily seen to be the identity matrix. In other words, the chain never changes state.

Example 87. Let $p_{1,2} = p_{2,1} = 1$. Clearly, $|1 - p_{1,2} - p_{2,1}| = 1$, and therefore (118) does not apply. The transition probability matrix $\P$ is given by:

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

It can be verified by induction that:

$$\P(n) = \P^n = \begin{cases} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} & \text{if } n \text{ is even} \\ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} & \text{if } n \text{ is odd} \end{cases}$$

This Markov chain has an interesting behavior. Starting in state 1 (or 2), we return to state 1 (or 2) after an even number of steps. Therefore, the time between visits to a given state exhibits a periodic behavior. Such a chain is called a periodic Markov chain (with period 2). (Formal definition is given below).

§35.INVERSE TRANSFORM METHOD

Assuming our computer can hand us, up on demand, independent and identically distributed (iid) copies of random variables that are uniformly distributed on interval $(0, 1)$, it is imperative that we be able to use these uniforms to generate random variables of any desired distribution (exponential, Bernoulli etc.). The first general method that we present is called the inverse transform method.

Let $F(x)$, $x \in \mathbf{R}$, denote any distribution function (continuous or not). Recall that $F : \mathbf{R} \to (0, 1)$ is thus a non-negative and non-decreasing (monotone) function that is continuous from the right and has left hand limits, with values in $[0, 1]$. Moreover, we have $F(+\infty) = 1$ and $F(-\infty) = 0$. Our objective is to generate (simulate) random variables X distributed as $F(x)$, that is, we want to simulate a random variable X such that

$$P(X \leq x) = F(x), \qquad x \in \mathbf{R}.$$

Define the **generalized inverse** of F,

$$F^{-1} : [0, 1] \to \mathbf{R},$$

via

$$F^{-1}(y) = \min\{x : F(x) \geq y\}, \qquad y \in [0, 1]. \tag{35.1}$$

If F is continuous, then F is invertible (since it is thus continuous and strictly increasing) in which case

$$F^{-1}(y) = \min\{x : F(x) = y\},$$

the ordinary inverse function and thus

$$F(F^{-1}(y)) = y \qquad \text{and} \qquad F^{-1}(F(x)) = x.$$

In general it holds that

$$F^{-1}(F(x)) \leq x \qquad \text{and} \qquad F(F^{-1}(y)) \geq y.$$

$F^{-1}(y)$ is a non-decreasing (monotone) function in y. This simple fact yields a simple method for simulating a random variable X distributed as F:

Theorem 35.1 (The Inverse Transform Method). Let $F(x)$, $x \in \mathbf{R}$, denote any distribution function (continuous or not). Let $F^{-1}(y)$, $y \in [0,1]$ denote the inverse function defined in (35.1). Define $X = F^{-1}(U)$, where U has the continuous uniform distribution over the interval $(0,1)$. Then X is distributed as F, that is,

$$P(X \le x) = F(x), \qquad x \in \mathbf{R}.$$

Proof: We have to show that

$$P(F^{-1}(U) \le x) = F(x), \qquad x \in \mathbf{R}.$$

First suppose that F is continuous. Then we will show that (equality of events)

$$\{F^{-1}(U) \le x\} = \{U \le F(x)\},$$

so that by taking probabilities (and letting $a = F(x)$ in $P(U \le a) = a$) yields the result:

$$P(F^{-1}(U) \le x) = P(U \le F(x)) = F(x).$$

To this end: $F(F^{-1}(y)) = y$ and so (by monotonicity of F) if $F^{-1}(U) \le x$, then $U = F(F^{-1}(U)) \le F(x)$, or $U \le F(x)$. Similarly $F^{-1}(F(x)) = x$ and so if $U \le F(x)$, then $F^{-1}(U) \le x$. We conclude equality of the two events as was to be shown. In the general (continuous or not) case, it is easily shown that

$$\{U < F(x)\} \subset \{F^{-1}(U) \le x\} \subset \{U \le F(x)\},$$

which yields the same result after taking probabilities (since $P(U = F(x)) = 0$ since U is a continuous random variable.)