2.3 Steady-State Distribution; Regular Markov Chains

It is sometimes necessary to be able to make *long-term* forecasts, meaning we want

$$\lim_{h\to\infty} P_h,$$

so we need to compute $\lim\limits_{h\to\infty} p_{ij}^{(h)}$.

Definition 2.13. *Let X be a Markov chain. The vector $\pi = [\pi_1, \ldots, \pi_n]$, consisting of the limiting probabilities*

$$\pi_k = \lim_{h\to\infty} P_h(k), k = 1, \ldots, n, \tag{2.13}$$

if it exists, is called a ***steady-state (stationary, limiting) distribution*** *of X.*

When this limit exists, it can be used as a forecast of the distribution of X after *many* transitions.

In order to find it, let us notice that

$$P_h P \;=\; (P_0 P^h) P \;=\; P_0 P^{h+1} \;=\; P_{h+1}.$$

Taking the limit as $h \to \infty$ on both sides, we get

$$\pi P \;=\; \pi. \tag{2.14}$$

System (2.14) is an $n \times n$ *singular* linear system (multiplication by a constant on each side leads to infinitely many solutions). However, since π must also be a *stochastic* matrix, the sum of its components must equal 1. Thus, we add one more condition,

$$\pi_1 + \pi_2 + \ldots \;=\; 1,$$

called the *normalizing* equation. If a solution of system (2.14) *exists*, then this extra condition will make it *unique*.

We state the following result, without proof.

Proposition 2.14. *The steady-state distribution of a homogeneous Markov chain X, $\pi = [\pi_1, \ldots, \pi_n]$, if it exists, is unique and is the solution of the $(n+1) \times n$ linear system*

$$\left\{ \begin{array}{lcl} \pi P & = & \pi \\ \sum\limits_k \pi_k & = & 1. \end{array} \right. \tag{2.15}$$

Example 2.15. Let us find the steady-state distribution of the Markov chain in Example 2.10 (Lecture 5). What is the weather forecast in Rainbow City for Christmas Day next year?

Solution. Recall that in Example 2.10 we had a homogeneous Markov chain with two states, (1-sunny, 2-rainy), the initial situation (on Monday) was 80% chance of rain, i.e.

$$P_0 = [0.2 \;\; 0.8]$$

and the transition probability matrix was

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}.$$

We write system (2.14). We have

$$[\pi_1 \;\; \pi_2]P = [\pi_1 \;\; \pi_2] \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} = \begin{bmatrix} 0.7\pi_1 + 0.4\pi_2 \\ 0.3\pi_1 + 0.6\pi_2 \end{bmatrix},$$

so system (2.14) becomes

$$\begin{cases} 0.7\pi_1 + 0.4\pi_2 = \pi_1 \\ 0.3\pi_1 + 0.6\pi_2 = \pi_2 \end{cases} \iff 0.3\pi_1 - 0.4\pi_2 = 0 \iff \pi_2 = \frac{3}{4}\pi_1.$$

We see that two equations in the system reduced to one. This will *always* happen, i.e., one equation will follow from the others, and this is because the system $\pi P = \pi$ is *singular*. It remains to use the normalizing equation, to get

$$\begin{cases} 3\pi_1 - 4\pi_2 = 0 \\ \pi_1 + \pi_2 = 1, \end{cases}$$

with solution

$$[\pi_1 \;\; \pi_2] = [4/7 \;\; 3/7].$$

Interpretation: in the long-run, in the future, $4/7 \approx 57\%$ of days are sunny and $3/7 \approx 43\%$ of days are rainy. Recall that the forecast for Wednesday was $53.8\%/46.2\%$ and for Friday, $56.84\%/43.16\%$, which is already getting close to the steady-state distribution.

Since Christmas Day next year is *many* steps from now, we use the steady-state distribution instead. So that would be the forecast for Christmas Day next year, too!

■

Remark 2.16.

1. Just as we did in the previous example, when we need to make predictions after a large number of steps, instead of the lengthy computation of P_h (i.e., P^h), it may be easier to try to find the steady-state distribution, π, directly.
2. If a steady-state distribution exists, then it can be shown that the matrix $P^{(h)} = P^h$ also has a limit, as $h \to \infty$, and the limiting matrix is given by

$$\mathbf{\Pi} \;=\; \lim_{h\to\infty} P^{(h)} \;=\; \begin{bmatrix} \pi \\ \vdots \\ \pi \end{bmatrix} \;=\; \begin{bmatrix} \pi_1 & \pi_2 & \dots & \pi_n \\ \vdots & \vdots & \dots & \vdots \\ \pi_1 & \pi_2 & \dots & \pi_n \end{bmatrix}.$$

3. Notice that π and $\mathbf{\Pi}$ *do not* depend on the initial state X_0. Actually, in the long run, the probabilities of transitioning from any state to a given state are the same, $p_{ik} = p_{jk},\ \forall i, j, k = \overline{1, n}$ (all the rows of $\mathbf{\Pi}$ coincide). Then, it is just a matter of "reaching" a certain state (from anywhere), rather than "transitioning" to it (from another state). That should, indeed, depend only on the pattern of changes, i.e. only on the transition probability matrix.
4. What is actually the "steady" state of a Markov chain? Suppose the system has reached its steady state, so that the current distribution of states is

$$P_t = \pi.$$

Then the system makes one more transition, and the distribution becomes

$$P_{t+1} = \pi P.$$

But $\pi P = \pi$ and thus,

$$P_t = P_{t+1}.$$

We see that in a steady state, transitions do not affect the distribution. A system may go from one state to another, but the *distribution* (the pdf) of states *does not change*. In this sense, it is *steady*.

Now, a natural question arises: does a steady-state distribution always exist? The answer is **no**! Here is a simple example:

Example 2.17. In a game of chess, a knight (în rom. "calul") can only move to a field of different color (white-to-black or black-to-white) at any time. Then the transition probability matrix of the color of its field is

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

For this matrix, a simple computation yields

$$P^2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I,$$

so,

$$\begin{aligned} P^{2k} &= I \text{ and} \\ P^{2k+1} &= P, \forall k \in \mathbb{N}. \end{aligned}$$

These are the *only* possible values. Thus,

$$\lim_{h \to \infty} P^h$$

does not exist and neither does

$$\lim_{h \to \infty} P_h.$$

This is a **periodic** Markov chain with period 2,

$$X_t = X_{t+2}.$$

Periodic Markov chains *do not* have a steady-state distribution.

There are other situations when steady-state probabilities cannot be found. So, when *does* a steady-state distribution exist? This is an ongoing research problem. We mention (without proof) one case, which is really easy to check, when such a distribution does exist.

Definition 2.18. *A Markov chain is called* ***regular*** *if there exists* $h \geq 0$*, such that*

$$p_{ij}^{(h)} > 0, \tag{2.16}$$

for all $i, j = 1, \ldots, n$.

This is saying that at some step h, $P^{(h)}$ has *only* non-zero entries, meaning that h-step transitions from any state to any state are possible.

Proposition 2.19. *Any regular Markov chain has a steady-state distribution.*

Remark 2.20. Regularity of Markov chains does not mean that *all* $p_{ij}^{(h)}$ should be positive, for *all* h. The transition probability matrix P, or some of its powers, may have some 0 entries, but there must exist some power h, for which $P^{(h)}$ has all non-zero entries.

Example 2.21. The Markov chain in Example 2.15 is regular because all transitions are possible for $h = 1$ already, and matrix P does not contain any zeros. Indeed, it has a steady-state distribution.

Example 2.22. A Markov chain with transition probability matrix

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0.9 & 0 & 0 & 0.1 \end{bmatrix}$$

is also regular. Matrix P contains zeros and so do P^2, P^3, P^4 and P^5. However, the 6-step transition probability matrix

$$P^{(6)} = \begin{bmatrix} .009 & .090 & .900 & .001 \\ .001 & .009 & .090 & .900 \\ .810 & .001 & .009 & .180 \\ .162 & .810 & .001 & .027 \end{bmatrix}$$

contains no zeros and shows regularity of this Markov chain.

In fact, computation of all P^h up to $h = 6$ is not even required in this case. Regularity can also be seen from the transition diagram in Figure 1. We can see that any state $j = 1, 2, 3, 4$ can be reached in 6 steps from any state $i = 1, 2, 3, 4$. Indeed, moving counterclockwise through this figure, we can reach state 4 from any state i in at most 3 steps. Then, we can reach any state j from state 4 again in at most 3 additional steps, for the total of at most 6 steps. If we can reach a state i from a state j in *fewer* than 6 steps, we just use the remaining steps circling around state 4. For example, state 2 is reached from state 1 in 6 steps as follows:

$$1 \to 2 \to 3 \to 4 \to 4 \to 1 \to 2.$$

Then, indeed all $p_{ij}^{(6)}$ are positive and the chain is regular. This goes to show that we don't have to *actually compute* all $p_{ij}^{(h)}$. We only need to verify that they are all positive for some h.

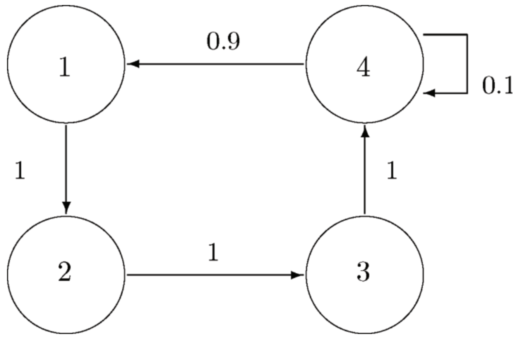

Fig. 1: Transition diagram for the regular Markov chain in Example 2.22

Absorbing states

If there exists a state i with $p_{ii} = 1$, then that Markov chain *cannot* be regular. There is no exit (no transition possible) from state i. Such a state is called an **absorbing state**. For example, state 4 in Figure 2(a) is absorbing, therefore, the Markov chain is irregular.

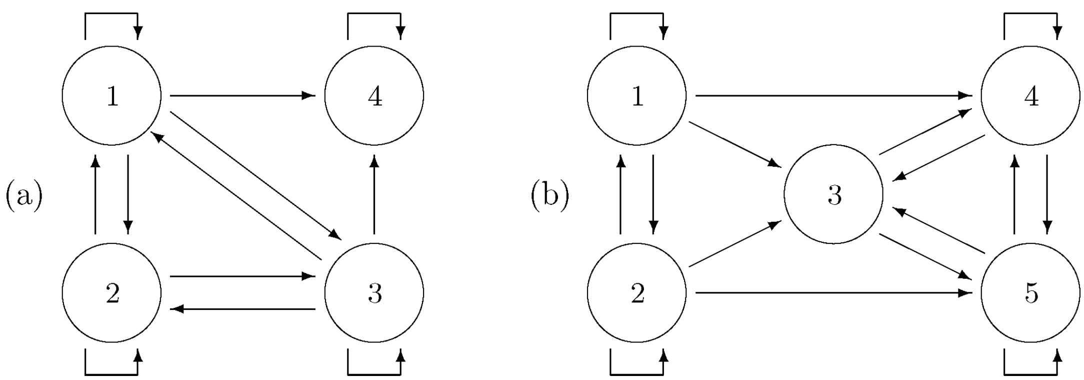

Fig. 2: Absorbing states and absorbing zones

There may be several absorbing states or an entire *absorbing zone*, from which the remaining states can never be reached. For example, states $3, 4$ and 5 in Figure 2(b) form an absorbing zone, some kind of a "Bermuda triangle". When this process finds itself in the set $\{3, 4, 5\}$, there is no route from there to the set $\{1, 2\}$. As a result, e.g. probability $p_{31}^{(h)}$ is 0 for *all* h.

However, notice that both Markov chains *do* have steady-state distributions. The first process will eventually reach state 4 and will stay there for good. Therefore, the limiting distribution of X_h is

$$\pi \;=\; \lim_{h\to\infty} P_h \;=\; [0\;\;0\;\;0\;\;1].$$

The second Markov chain will eventually leave states 1 and 2 for good, thus its limiting (steady-state) distribution has the form

$$\pi \;=\; [0\;\;0\;\;\pi_3\;\;\pi_4\;\;\pi_5].$$

This goes to show that the converse of Proposition 2.19 is *not true*, there are irregular Markov chains that have a steady-state distribution.

Remark 2.23. The study of Markov chains gives us an important method of analyzing rather complicated stochastic systems. Once the Markov property of a process is established, it only remains to find its one-step transition probabilities. Then, the steady-state distribution can be computed, and thus, we obtain the distribution of the process *at any time*, after a sufficient number of transitions. This methodology will be our main working tool in the next chapter, when we study queuing systems and evaluate their performance.

3 Counting Processes

A special case of stochastic processes are the ones where one needs to count the occurrences of some types of events over time. These are described by *counting processes*.

Definition 3.1. *A* ***counting process*** $X(t), t \geq 0$*, is a stochastic process that represents the number of items counted by the time* t*.*

Counting processes deal with the number of occurrences of something over time, such as customers arriving at a supermarket, completed tasks, transmitted messages, detected errors, scored goals, number of job arrivals to a queue, holding times (in renewal processes), etc.

In general, we refer to the occurrence of each event that is being counted as an "arrival". As time passes, one can count additional items. Therefore, sample paths (values) of a counting process are always *non-decreasing, non-negative integers* $\{0, 1, \ldots\}$.

Thus, **all** counting processes are **discrete-state** stochastic processes. They can be discrete-time or continuous-time.

Example 3.2. Figure 3 shows sample paths of two counting processes, $X(t)$ being the number of transmitted e-mails by the time t and $A(t)$ being the number of transmitted attachments. According to the graphs, e-mails were transmitted at times $t = 8, 22, 30, 32, 35, 40, 41, 50, 52$ and 57 min. The e-mail counting process $X(t)$ increments by 1 at each of these times. Only 3 of these e-mails contained attachments. One attachment was sent at time $t = 8$, five more at $t = 35$, making the total of $A(35) = 6$, and two more attachments at $t = 50$, making the total of $A(50) = 8$.

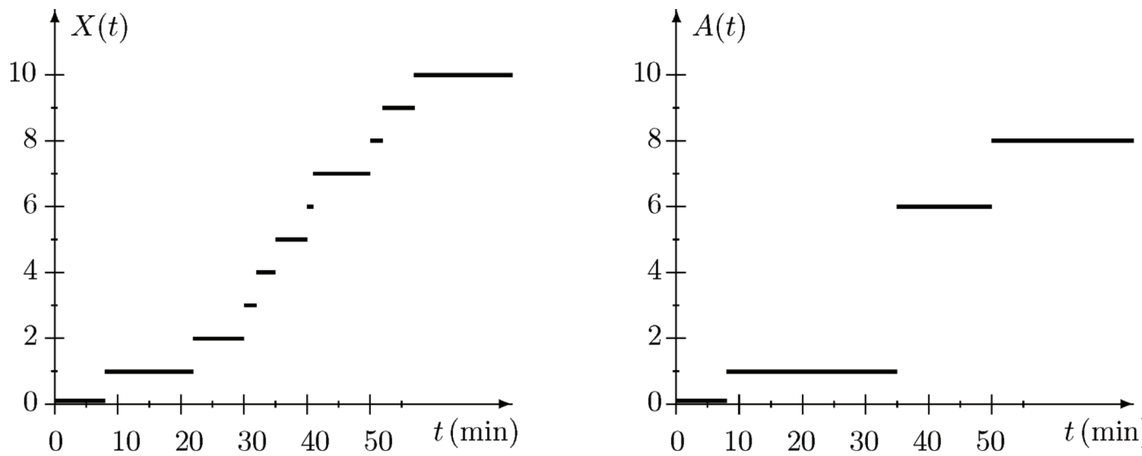

Fig. 3: Counting processes in Example 3.2

Next, we consider the most widely used examples, Binomial (discrete-time) and Poisson (continuous-time) counting processes.

3.1 Binomial Counting Process

Consider a sequence of Bernoulli trials with probability of success p and count the number of "successes".

Definition 3.3. *A* ***Binomial counting process*** $X(n)$ *is the number of successes in* n *Bernoulli trials,* $n = 0, 1, \ldots$.

Remark 3.4.
1. Obviously, a Binomial process $X(n)$ is a discrete-state, discrete-time stochastic process, "time" being measured discreetly, by the number of trials, n.
2. The pdf of $X(n)$ is Binomial $B(n, p)$ at any time n (see Figure 4). Recall that

$$E(X(n)) = np.$$

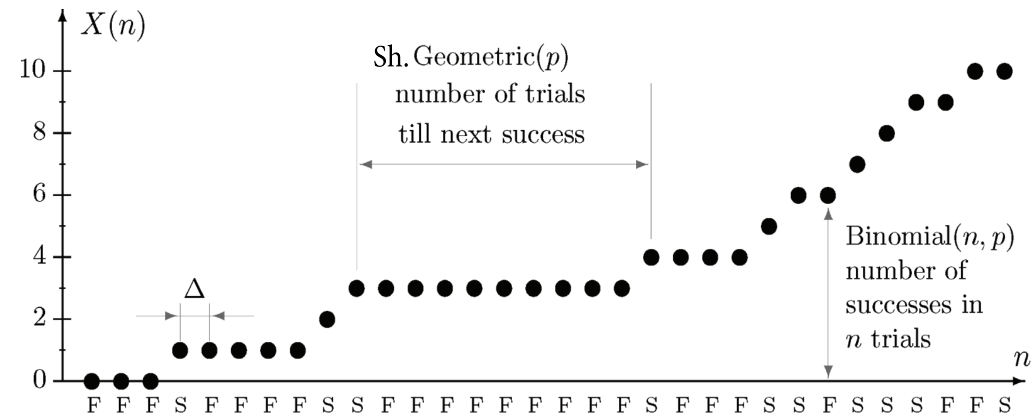

Fig. 4: Binomial process sample path (S = success, F = failure)

3. The number of trials between two consecutive successes, Y, is the *number of trials needed to get the next (first) success*, so it has a $SGeo(p)$ pdf (see Figure 4). Recall that

$$E(Y) = \frac{1}{p},\ V(Y) = \frac{q}{p^2}.$$

Relation to real time, frames

It is important to make the distinction between real time and the "time" variable n ("time" as in a stochastic process). Variable n is not measured in time units, it measures the number of trials.

Suppose that Bernoulli trials occur at equal time intervals, say every Δ seconds (see Figure 4). That means that n trials occur during time $t = n\Delta$. The value of the process at time t has Binomial pdf with parameters $n = \dfrac{t}{\Delta}$ and p. Then the expected number of successes *during t seconds* is

$$E(X(n)) \;=\; E\left(X\left(\frac{t}{\Delta}\right)\right) \;=\; np \;=\; \frac{t}{\Delta}p \;=\; t\frac{p}{\Delta},$$

so the expected number of successes *per second* is

$$\lambda \;=\; \frac{p}{\Delta}.$$

Definition 3.5.

- *The quantity* $\lambda = \dfrac{p}{\Delta}$ *is called the* ***arrival rate****, i.e. the average number of successes per one unit of time.*
- *The quantity* Δ *is called a* ***frame****, i.e the time interval of each Bernoulli trial.*
- *The* ***interarrival time*** *is the time between successes.*

We can now rephrase:

– p is the probability of *arrival* (success) during one *frame* (trial),

– $n = \dfrac{t}{\Delta}$ is the number of *frames* during time t,

– $X\left(\dfrac{t}{\Delta}\right)$ is the *number of arrivals* by time t.

The concepts of *arrival rate* and *interarrival time* deal with modeling arrival of jobs in discrete-time queuing systems by Binomial counting processes. The key assumption in such models is that no more than 1 arrival can occur during each Δ-second frame (otherwise, a smaller Δ should be considered), so each frame is a Bernoulli trial.

The interarrival period, Y, measured in number of frames, has a $SGeo(p)$ pdf (as mentioned earlier). Since each frame takes Δ seconds, the interarrival time is

$$T = \Delta Y,$$

a *rescaled* $SGeo(p)$ variable, whose expected value and variance are given by

$$\begin{aligned} E(T) &= \Delta E(Y) = \Delta \frac{1}{p} = \frac{1}{\lambda}, \\ V(T) &= \Delta^2 V(Y) = \Delta^2 \frac{q}{p^2} = \frac{1-p}{\lambda^2}. \end{aligned} \tag{3.1}$$

Example 3.6. Messages arrive at a communications center at the rate of 6 messages per minute. Assume arrivals of messages are modeled by a Binomial counting process.
a) What frame size should be used to guarantee that the probability of a message arriving during each frame is 0.1?
b) Using the chosen frames, find the probability of no messages arriving during the next 1 minute.

c) Compute the probability of more than 35 messages arriving during the next 6 minutes.

d) Find the probability of more than 350 messages arriving during the next hour.

e) What is the average interarrival time and its standard deviation?

f) Compute the probability that the next message does not arrive during the next 20 seconds.

Solution.

a) We have $\lambda = 6$ / min. and $p = 0.1$. Thus,

$$\Delta = \frac{p}{\lambda} = \frac{1}{60}\text{ min.} = 1\text{ sec.}$$

b) So $\Delta = 1$ sec. In $t = 1$ minute $= 60$ seconds, there are $n = \dfrac{t}{\Delta} = 60$ frames. The number of messages arriving during 1 minute (i.e. 60 frames), $X(60)$, has a Binomial distribution with parameters $n = 60$ and $p = 0.1$. So the desired probability is

$$\begin{aligned} P\big(X(60) = 0\big) &= \mathrm{pdf}_{X(60)}(0) \\ &= \mathrm{binopdf}(0, 60, 0.1) \\ &= 0.0018. \end{aligned}$$

c) Similarly, in $t = 6$ minutes $= 360$ seconds, there are $n = \dfrac{t}{\Delta} = 360$ frames. So, the number of messages arriving during the next 6 minutes, $X(360)$, has Binomial distribution with parameters $n = 360$ and $p = 0.1$. Then the probability of more than 35 messages arriving during the next 6 minutes is

$$\begin{aligned} P\big(X(360) > 35\big) &= 1 - P\big(X(360) \le 35\big) \\ &= 1 - \mathrm{cdf}_{X(360)}(35) \\ &= 1 - \mathrm{binocdf}(35, 360, 0.1) \\ &= 0.5257. \end{aligned}$$

d) Again, in $t = 1$ hour $= 3600$ seconds, there are $n = \dfrac{t}{\Delta} = 3600$ frames. Thus, the number of messages arriving during one hour, $X(3600)$, has Binomial distribution with parameters $n = 3600$

and $p = 0.1$. Then the probability of more than 350 messages arriving during the next hour is

$$\begin{aligned} P\big(X(3600) > 350\big) &= 1 - P\big(X(3600) \le 350\big) \\ &= 1 - \mathrm{cdf}_{X(3600)}(350) \\ &= 1 - \mathrm{binocdf}(350, 3600, 0.1) \\ &= 0.6993. \end{aligned}$$

<u>**Notice**</u> that "more than 35 messages in 6 minutes" is **not** the same as "more than 350 messages in 60 minutes"!! These are *random* variables ...

e) By (3.1), we have

$$\begin{aligned} E(T) &= \frac{1}{\lambda} = \frac{1}{6}\text{ minutes} = 10\text{ seconds}, \\ \mathrm{Std}(T) &= \sqrt{V(T)} = \sqrt{\frac{1-p}{\lambda^2}} = \sqrt{0.0250}\text{ minutes} \approx 9.5\text{ seconds}. \end{aligned}$$

f) Recall that the interarrival time $T = \Delta Y$, where Y has a $SGeo(p)$ distribution and, hence, $Y - 1$ has a $Geo(p)$ pdf. The next message does not arrive during the next 20 seconds, if $T > 20$. So,

$$\begin{aligned} P(T > 20) &= P(\Delta Y > 20) = P\left(Y > 20/\Delta\right) = P\left(Y > 20\right) \\ &= 1 - P(Y \le 20) = 1 - P(Y - 1 \le 19) \\ &= 1 - \mathrm{cdf}_{Y-1}(19) = 1 - \mathrm{geocdf}(19, 0.1) \\ &= 0.1216. \end{aligned}$$

Alternatively, this is also the probability of 0 arrivals during the next $t = 20$ seconds, i.e. during $n = \dfrac{t}{\Delta} = 20$ frames. The number of messages arriving during the next 20 seconds, $X(20)$, has a Binomial distribution with parameters $n = 20$ and $p = 0.1$. Thus, the probability that no messages arrive during the next 20 seconds is

$$P\big(X(20) = 0\big) = \mathrm{pdf}_{X(20)}(0) = \mathrm{binopdf}(0, 20, 0.1) = 0.1216.$$

■

Markov property of Binomial counting processes

It is clear that the number of successes in n trials depends *only* on the number of successes in $n-1$ trials (not on previous values $n-2, n-3, \dots$), so a Binomial process has the Markov property. Thus, it is a **Markov chain**.

Let us find the transition probability matrix. At each trial (i.e. during each frame), the number of successes $X(n)$ either increases by 1 (in case of success), or stays the same (in case of failure). Then,

$$p_{ij} = \begin{cases} p, & j = i+1 \\ q = 1-p, & j = i \\ 0, & \text{otherwise} \end{cases}. \tag{3.2}$$

Obviously, transition probabilities are constant over time and independent of past values of $X(n)$. Hence, $X(n)$ is a **homogeneous Markov chain** with transition probability matrix given by

$$P = \begin{bmatrix} 1-p & p & 0 & \dots & 0 & \dots \\ 0 & 1-p & p & \dots & 0 & \dots \\ 0 & 0 & 1-p & \dots & 0 & \dots \\ \vdots & \vdots & \vdots & & \vdots & \end{bmatrix} \tag{3.3}$$

and transition diagram depicted in Figure 5.

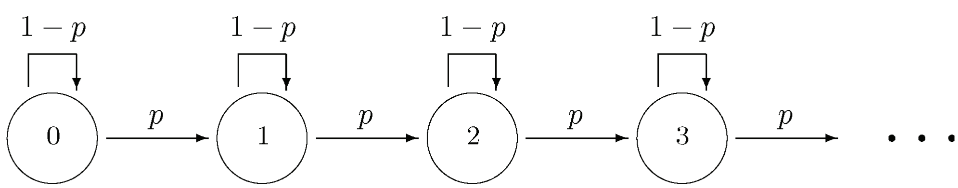

Fig. 5: Transition diagram for a Binomial counting process

Notice that it is an *irregular* Markov chain. Since $X(n)$ is non-decreasing, e.g. $p_{10}^{(h)} = 0$, for all $h \geq 0$ (once we have a success, the number of successes will *never* go back to 0). A Binomial counting process *does not* have a steady-state distribution.

Another interesting fact: the h-step transition probabilities simply form a Binomial distribution.

Indeed, $p_{ij}^{(h)}$ is the probability of going from i to j successes in h transitions, i.e.,

$$
\begin{aligned}
p_{ij}^{(h)} &= P\big((j-i) \text{ successes in } h \text{ trials}\big) \\
&= \begin{cases} C_h^{j-i} p^{j-i} q^{h-j+i}, & 0 \le j-i \le h \\ 0, & \text{otherwise} \end{cases}.
\end{aligned}
$$

Simulation of Binomial counting processes

This is straightforward, a sequence of Bernoulli trials, where we count the number of successes.

Algorithm 3.7.
1. Given: $N_B =$ sample path length of the Binomial counting process,
2. Generate $U \in U(0,1)$, let $Y = (U < p)$, let $X(1) = Y$.
3. At each time t, let $Y = (U < p)$, let $X(t) = X(t-1) + Y$.
4. Return to step 3 until length N_B is achieved.