F. Borgonovo

Short notes on Probability Theory

V. September 2017

Contents

Chapter 1

Probabilities

1.1 Finite spaces: Definitions and axioms

Let $\mathcal{E}$ be an experiment which provides a finite number of outcomes $\alpha_1, \alpha_2, \ldots, \alpha_n$. We state the following definitions:

- **Trial** is the execution of $\mathcal{E}$ which leads to a outcome, or sample, α and only one.
- **Space or stochastic universe** associated with the experiment $\mathcal{E}$ is the set $S = \alpha_1, \alpha_2, \ldots . \alpha_n$ of all possible outcomes of $\mathcal{E}$.
- **Event** is any set A of outcomes and what is a any subset of S.
- An **Elementary Event or Sample Event** is a set $E \subset S$ with a single outcome.
- A **Certain Event** is an event that coincides with S.
- The **Impossible Event** is the empty set $\emptyset$.
- Events can be combined with operations in use in Set theory, yielding events such as **union or sum**, **conjunction or product or intersection**, **complement**, difference.
- **We say that in a trial event A occurs if the outcome of the trial belongs to A.**

From the above it follows that:

- the certain event always occurs;
- the impossible event never occurs;
- a union event occurs if at least one of the component events occurs;
- a joint event occurs if all the components events occur simultaneously;

- disjoint events can not occur simultaneously, and for this reasons they are called mutually exclusive; Such events are, for example, sample events complementary events.

On the concept of probability. The probability $P(A)$ of an event A is a tool that reveals its usefulness in measuring the degree of occurrence of event A during one or more experiments where A may occur, eg the occurrence of "head" when a coin is tossed. Its meaning can not be explained "a priory", exactly as it happens for other concepts of mathematics or geometry, eg, a straight line. On the other side, the Probability Theory is not based on the "meaning" of probability. It deals with rules that allow to evaluate the probability of some events given the probabilities of some other events. In other words, the Probability Theory deals with the rules for manipulating probabilities. Here we define the probability of an event as a kind of "arbitrary" real function on the space of the events that must obey to some simple rules, mandatory for the internal coherence of the theory. Nevertheless, we will see later on that this theory let us attach a meaning to probability, and a consequent "measure", allowing direct applications to the real word.

Definition
The probability [1] $P(A)$ *of an event* $A \subset S$ *is a"* **measure** *"defined on* S *so as to satisfy the following:*

Axiom I*:* $P(A)$ *is a nonnegative real number associated with the event.*

$$P(A) \geq 0 \tag{1.1}$$

Axiom II*: the probability of the certain event is one.*

$$P(S) = 1 \tag{1.2}$$

Axiom III*: if* A *e* B *are disjoint events*

$$P(A + B) = P(A) + P(B) \tag{1.3}$$

We have the following corollaries

Corollary 1

$$P(A) = 1 - P(\bar{A}) \leq 1 \tag{1.4}$$

In fact we have $A + \bar{A} = S$ e $A\bar{A} = \phi$ and the thesis follows from (1.3).

Corollary 2

$$P(\phi) = 0 \tag{1.5}$$

In fact we have $\phi = \bar{S}$ and the thesis follows from (1.4).

[1] We use the short term "probability of A "instead of the more correct "probability that the outcome of a trial verifies $A \subset S$ "

Corollary 3: *If $B \subset A$, then*

$$P(B) \leq P(A) \tag{1.6}$$

In fact we have $A = B + (A\bar{B})$ where B and $A\bar{B}$ are disjoint events, and the thesis follows from (1.3) e (1.1).

Corollary 4: *If $A_1, A_2, \ldots, A_n$ are disjoint events, and $A = A_1 + A_2 + \ldots + A_n$, then we have*

$$P(A) = P(A_1) + P(A_2) + \ldots + P(A_n). \tag{1.7}$$

The thesis follows from repeated application of (1.3).

The definitions and corollaries above allow to manipulate spaces where the outcomes of the experiment are finite in number. For example, assume that the outcomes are represented by the face value that appears when a dice is rolled. We have six outcomes and the probabilities of any event, such as event $\{$*an even face appears*$\}$ can be evaluated if we know the probabilities of all elementary events represented by the outcomes.

Definition
We say that an experiment $\mathcal{E}$ (or probability space S) is completely described by the probabilistic point of view when, for each elementary event (outcome) E_i it is given, or it is possible to obtain the corresponding probability

$$p_i = P(E_i)$$

In this case Corollary 4 allows to derive the probability of any kind of event A as the sum of the probabilities of the elementary events that compose A.

Corollary 5: *If all the n_S elementary events E_i are equally probable, the space S is called* **uniform** *and the probability of an event A composed of r_A elementary events is:*

$$P(A) = \frac{r_A}{n_S} \tag{1.8}$$

In fact, being by hypothesis $P(e_i) = p$, $(i = 1, 2, \ldots)$ and being the elementary events disjoint by definition, applying the (1.7) yields

$$P(A) = \sum_i P(E_i) = p\ r_A \quad \text{(for every } E_i \subset A)$$

and from (1.1) and (1.7)

$$1 = P(S) = \sum_i P(E_i) = p\ n_S \quad \text{(for every } E_i \subset S)$$

Subdividing the above expressions we get (1.8).♣

Resuming the example of a dice rolled, in the case that all faces are equally probable, the preceding Corollary shows, for example, $P(\textit{an even face appears}) = 0.5$. However, we can also consistently assume that faces $1, 2, 3, 4, 5, 6$ occur with probabilities $0, 0.5, 0.1, 0.2, 0.1, 0.1$, (note that the sum must equal 1 by Axiom II. In this case by Axiom III repeatedly applied to events $2, 4, 6$, provides $P(\textit{an even face appears}) = 0.8$. This latter example is not "wrong", as some could assume based on experience. In fact, the probabilistic description we have used is not dictated by the theory, and both results above are consistent. If we want to apply these results to the real world, then we must use a probabilistic description the fits the real world. In some cases it may be useful to assume the first model (all outcomes are equally probable), and we say the dice is "fair"; in some other cases may be more convenient assume a different model, for example when we suspect that the dice is not fair. To refer to the real world we need a *measure* of the fairness of the dice under consideration, i.e., we need to be able to measure the probabilities of its outcomes.

Here we anticipate a fundamental result of the Theory, the **Law of Large Numbers**, that is very simple in its meaning, but is complex to expose in a correct manner. So, for the moment, we state it in some approximated form, very useful for undertaking measures. Let denote by N the number of times an experiment is performed, and let N_A the number of times event A is verified. Then

$$P(A) \approx \lim_{N\to\infty} \frac{N_A}{N} \tag{1.9}$$

The approximation becomes more and more precise as N grows, however, given the finiteness of N, the exact measure can never be accomplished in practice.

Many of the probability calculations we will see later on are related to uniform spaces and, therefore, probability evaluations are based on the preceding Corollary and are attained by counting techniques (such as combinatorial calculus) in order to get r_A and r_n.

1.1.1 Simulation

Probabilistic experiments can be built that are able to provide a space of outcomes S with any probability description, starting by a simple *basic experiment*, capable to provide outcomes with the *same probability* (uniform). We assume that this experiment can be represented by the extraction of a ball from an urn that contains n balls numbered from 1 to n. The probability of each outcome is then $1/n$.

Example (1.10)
We want to build an experiment that represents the outcomes of an unfair coin tossing that shows head with probability 0.6.

The space of the basic experiment must be divided into two events, say H (head) and T (tail), with probabilities $p_H = 0.6$ and $p_T = 0.4$ respectively. This can be done, for example, using a basic experiment with 10 balls and by defining event H as the set of balls whose number is between 1 to 6. If the drawn ball belongs to this set we say H has occurred, otherwise T has occurred. Note that the basic experiment can be performed also with $n = 5$, and H is defined as the drawing of a ball whose number is between 1 to 3.

As the probability of H is required with more digits, then the number of balls n in the basic experiment must be increased. For example, if we want $p_H = 0.634$, we must choose $n = 1000$, and

H is defined as the drawing of a ball whose number is between 1 to 634. This shows that if p_H is approaching a real number, then n should go to infinity.♣

In the same way as above we can define an experiment with multiple outcomes such as the throwing of a dice. In this case the state space of the basic experiment must be divided, uniformly or not, into six events representing the outcomes of the different faces of the dices.

The basic experiment we referred to can be provided also by many software tools. There are software routines that provides the number of the ball drawn from the urn. The maximum number can be very high, such as 10^{16} or more, in order to allow for fine grain probabilities. However, software can not provide random experiments, since computers are causal machines and results are always predictable. Nevertheless, those software uses algorithms such that the numbers provided mimic very well those provided by the urn extraction.

The probabilities of events attained as explained above can then be measured by using the law of large numbers explained above (1.9).

1.2 Uniform and finite spaces

The experiments that we consider in this section are modeled with the *urn model* where k of objects (elements) are drawn among the the n object in the urn. Here we assume that all outcomes consisting of all the groups that can be formed with k of n objects are equally probable, where groups differ in at least one element or in the order they appear in the group. If objects are drawn together, or one by one with no replacement, the number of such groups is

$$(n)_k = n(n-1)\dots(n-k+1),$$

otherwise, if the objects are drawn one by one with replacement, the number of such groups is

$$(n)_k = n^k$$

and probabilities can be evaluated via (1.8).

Example (1.11)

In an urn there are ten objects representing the ten digits $0, 1, \dots,$ 9. Evaluate the probability that, upon drawing of 3 elements, the three digits form the event

$A = \{$*number 567*$\}$

$B = \{$*number with three consecutive increasing digits.*$\}$.

We have $n_S = (10)_3 = 10 \cdot 9 \cdot 8 = 720$

$$r_A = 1 \qquad P(A) = \frac{1}{720}$$

$$r_B = 8 \qquad P(B) = \frac{8}{720}$$

Example (1.12)

From the Urn of the previous example the three drawings are occur with the replacement of single element previously extracted. Evaluate the probability of events A and B of the previous example and of event $C = \{$number with all equal digits. $\}$

$$n_S = 10^3$$

$$r_A = 1 \qquad P(A) = \frac{1}{1000}$$

$$r_B = 8 \qquad P(B) = \frac{8}{1000}$$

$$r_C = 10 \qquad P(C) = \frac{10}{1000}$$

Example (1.13)

Assuming that people have equal probability to be born any day of the year, evaluate the minimum number of people k you need to draw so that the probability of having at least two people born on the same day exceeds 0.5.

The experiment is similar to the extraction with replacement of k numbers out of an urn that contains 365 objects, each representing a different day of the year. Denoted $D_k = \{$ *extraction k objects all different* $\}$ we have

$$P(D_k) = \frac{(365)_k}{(365)^k} \qquad k \leq 365$$

The probability to set is

$$P(\overline{D}_k) = 1 - P(D_k) = 1 - \frac{365!}{(365-k)!\; 365^k}$$

with

$$P(\overline{D}_{10}) = 0,166 \ldots P(\overline{D}_{22}) = 0,47576 \ldots P(\overline{D}_{23}) = 0,5072 \ldots P(\overline{D}_{30}) = 0,7062 \ldots$$

The lowest k that provides $P(\overline{D}_k) > 0.5$ is then $k = 23$.

1.3 Union of non-disjoint events

Theorem: (1.14)

Given events A e $B \subseteq S$ the following relation holds

$$P(A+B) = P(A) + P(B) - P(AB) \tag{1.15}$$

The proof is attained by writing event $A + B$ as union of two disjoint events (see Figure 1.1):

$$A - AB \text{ (horizontal dash lines);} \quad AB \text{ (grid);} \quad B - AB \text{ (vertical dash lines)}$$

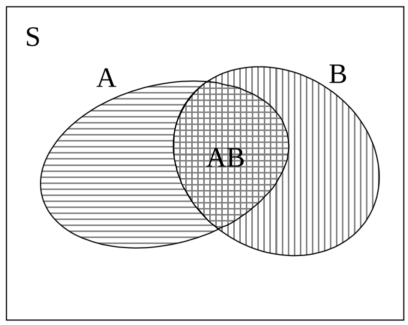


Figure 1.1:

From Corollary 4 we have

$$P(A + B) = P(A - AB) + P(AB) + P(B - AB),$$

observing that

$$P(A - AB) = P(A) - P(AB),$$
$$P(B - AB) = P(B) - P(AB),$$

and substituting, we get (1.15).♣

If A e B are disjoint events, (1.15) reduces to (1.3).

Example (1.16)
On a throw of the dice, evaluate the probability that a number appears that is even or less than 3.

Denote the event "even number" as A and the event "less than three" as B we have

$$P(D) = P(A + B) = P(A) + P(B) - P(AB),$$

We evaluate $P(A), P(B)$, and $P(AB)$ with the counting process and we find

$$P(A) = \frac{3}{6}, \qquad P(B) = \frac{2}{6}, \qquad P(AB) = \frac{1}{6}.$$

Substituting we get

$$P(D) = \frac{1}{3} + \frac{1}{2} - \frac{1}{6} = \frac{4}{6}.$$

1.4 Conditional spaces, events and probabilities

Assume an experiment $\mathcal{E}$, a space S and a probability measure $P(\cdot)$.

We want to evaluate the probability that the outcome α of a trial of $\mathcal{E}$ verifies A ($\alpha \in A$) knowing that event ($\alpha \in M$), with $P(M) > 0$, occurs.

Obviously, knowing (for example, because we are told) that the outcome verifies M gives some information with respect to the occurrence of a particular α (for example, all α that do not belong to M are excluded). In these conditions the probability of the occurrence of A is no longer the original one $P(A)$, which can be referred to as "a priori".

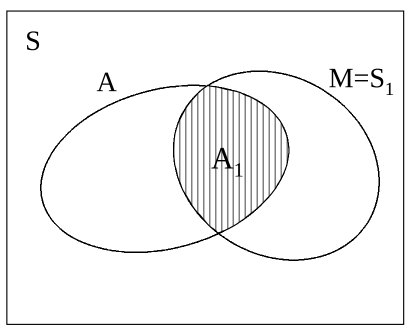


Figure 1.2:

We can formalize this concept by deriving from experiment $\mathcal{E}$ an experiment $\mathcal{E}_1$ (conditional experiment) the outcomes of which, α_1, are those of $\mathcal{E}$ that belong to M. Other outcomes are discarded and $\mathcal{E}$ is repeated. In this way the space (conditional) associated with $\mathcal{E}_1$ is $S_1 \equiv M$ and events A_1 of S_1 are called "conditional events".

Obviously, the probabilistic description $P_1(\cdot)$ in S_1 "must" be linked to description $P(\cdot)$ in S, i.e. the probability of the conditional event A_1 must be a function of the probability of the corresponding event AM (unconditional) in S:

$$P_1(A_1) = g[p(AM)].$$

Given A_1 e B_1, disjoint events, Axiom III says that

$$P_1(A_1 + B_1) = P_1(A_1) + P_1(B_1),$$

that is,

$$g[P(AM + BM)] = g[P(AM)] + g[P(BM)],$$

which implies that $g(\cdot)$ is a linear function of the type

$$g[P] = \alpha P + \beta,$$

α and β being constant values.

Axiom IIIa implies $\beta = 0$, whereas Axiom II forces

$$g[P(M)] = 1,$$

i.e.

$$\alpha = \frac{1}{P(M)}.$$

Then we have

$$P_1(A_1) = \frac{P(AM)}{P(M)}$$

However, instead of making use of formalization $\mathcal{E}_1$ and A_1, one prefer to write

$$P(A/M) = \frac{P(AM)}{P(M)}. \tag{1.17}$$

Relation (1.17) is meaningful only if $P(M) \neq 0$ and not only excludes $M = \phi$, but also the case where M is a elementary event in a continuous space. Such a conditional probability is defined later.

Example (1.18)

Evaluate the probability that the outcome of the roll of a dice is 2 knowing that the result is even.

We have

$$S = \{1,2,3,4,5,6\} \qquad S_1 = \{2,4,6\}$$

and from (1.17)

$$P(2/\text{ even}) = P_1(2) = \frac{1}{3}$$

1.4.1 Simulation of conditional events

Conditional events $\{A|M\}$ can be simulated by repeating the basic experiment as long as event M does not appear, and stopping when M appears. This provides an outcome that surely belongs to M.

1.5 Total Probability

It is often easy to determine the probability of an event A conditioned by other events M_i. In this case the probability of A can be determined as a function of the conditional probabilities by resorting to the so-called:

Theorem: Total Probability (1.19)
Given $M_1, M_2, \ldots M_n$ disjoint events such that $M_1 + M_2 + \ldots + M_n = S$ (really the less stringent condition $M_1 + M_2 \ldots M_N \supset A$ is sufficient), we have

$$P(A) = \sum_{i=1}^{n} P(A/M_i)P(M_i). \tag{1.20}$$

In fact, since events AM_i are disjoint, and their union provides A, we can write

$$P(A) = \sum_{i=1}^{n} P(AM_i), \tag{1.21}$$

and using the relation

$$P(AM_i) = P(A/M_i)P(M_i), \tag{1.22}$$

attained by reversing (1.17), we get (1.20). ♣

Example (1.23)
A box contains three types of objects, some of which are defective, in these proportions

type A - 2500 of which 10% defective
type B - 500 of which 40% defective
type C - 1000 of which 30% defective

If we draw an object at random, what is the probability $P(D)$ that, drawing an object, this is found to be defective?

Conditions are met for the validity of (1.20). The probability to draw an object of type A, B, C are respectively

$$P(A) = \frac{2500}{4000} = \frac{5}{8}; \qquad P(B) = \frac{500}{4000} = \frac{1}{8}; \qquad P(C) = \frac{1000}{4000} = \frac{2}{8}.$$

Then we have

$$P(D/A) = \frac{10}{100}; \qquad P(D/B) = \frac{40}{100}; \qquad P(D/C) = \frac{30}{100};$$

and, finally, from (1.20)

$$P(D) = P(D/A)P(A) + P(D/B)P(B) + P(D/C)P(C) = \frac{3}{16}$$

Obviously, in this case the probability can also be obtained as ratio between the number of defective items and the total number of objects:

$$P(D) = \frac{250 + 200 + 300}{4000} = \frac{3}{16}$$

However, in less fortunate cases this direct approach is not allowed.

Example (1.24)

A box contains n pieces of paper, each one reporting a number arbitrarily determined. At first r pieces are drawn and their maximum value m_r is observed. Further drawings are performed until a value m is observed such as $m > M_r$. You win if $m = M$, where M is the maximum value among those reported on the n pieces of paper. We want to evaluate the probability $P(V)$ to win, and the value of $r = r_m$ at which this probability is maximum.

If one assumes the position of the maximum equally likely, the probability that M is in position k is

$$P(M \text{ in } k) = \frac{1}{n} \tag{1.25}$$

The probability to win, with M in k, is zero if $k \leq\ r$. For $k > r$ you win if the maximum m_{k-1} among the first $k-1$ pieces of paper lies within the first r, and this happens with probability

$$P_r(V/M \text{ in } k > r) = \frac{r}{k-1}. \tag{1.26}$$

By the Total Probability Theorem (1.20) we have:

$$P_r(V) = \sum_{k=r+1}^{n} \frac{r}{k-1} \frac{1}{n} = \frac{r}{n} \sum_{k=r}^{n-1} \frac{1}{k}.$$

By the preceding result we can write

$$P_r(V) - P_{r+1}(V) = \frac{1}{n}(1 - \frac{1}{n-1} - \frac{1}{n-2} - \ldots - \frac{1}{r+1})$$

When r varies from $n-1$ to 1 the above expression changes from positive to negative and the sought value r_m is the smallest value of r for which the expression is maintained positive. For n and r large enough you can derive an approximate result by substituting, into the expression of $P_r(V)$, the sum with an integral obtaining

$$P_r(V) \simeq \frac{r}{n} \int_r^n \frac{1}{x} dx = \frac{r}{n} \ln \frac{n}{r}.$$

The above expression presents a maximum at

$$\frac{r}{n} = e^{-1} = 0,3675\ldots$$

where it is

$$P(V) = e^{-1} = 0,3675\ldots$$

In Table below exact values of r_m e $P(V)$ are reported for some n.

n	R_m	$P(V)$
3	1	0,5
4	1	$0,458\overline{3}$
5	2	$0,4\overline{3}$
7	2	$0,4142\ldots$
10	3	$0,3986\ldots$
50	18	$0,3742\ldots$
100	37	$0,3710\ldots$
1000	368	$0,3681\ldots$

1.6 Bayes' Formula

If we use (1.17) two times in a direct and a reverse way we get

$$P(M/A) = \frac{P(AM)}{P(A)} = P(A/M)\frac{P(M)}{P(A)}. \tag{1.27}$$

Referring to (1.20) the above expression can be re-written as

$$P(M_k/A) = \frac{P(A/M_k)P(M_k)}{\sum_{j=1}^{n} P(A/M_j)P(M_j)}. \tag{1.28}$$

Both formulas above are referred to as the Bayes' Theorem. In particular, the (1.28) is called "Bayes' rule for the a posteriori probability, "when the observation of the event A comes "after" (that is, knowing) the occurrence of M_i. In this case, $P(M_i)$ are called "a priori probabilities" and $P(M_i/A)$ "a posteriori probabilities".

Example (1.29)
An object drawn at random from the box in Example 1.23 is found to be defective. Evaluate the probabilities that it is of type A, B and C respectively.

Using Bayes' formula and the preceding results we have

$$P(A/D) = \frac{P(D/A)P(A)}{P(D)} = \frac{10}{30}; \qquad P(B/D) = \frac{P(D/B)P(B)}{P(D)} = \frac{8}{30}.$$

Similarly we have

$$P(C/D) = \frac{12}{30}.$$

Note that while "a priori" the most likely type of object is A, after observing that the object is defective the most likely type of object is C.

Example (1.30)
With reference to Example 1.24, let compute the probability that the maximum M is in position $r = k$ knowing that those who played won.

We use (1.28) where A represents victory and M_k represents maximum in position k. At denominator we have the probability to win already evaluated, while at numerator we have respectively (1.25) e (1.26). We have:

$$P(M_k/V) = \frac{C}{k-1} \qquad\qquad k > r$$

and zero elsewhere, being C the normalization constant.

You can see how the knowledge of a preceding victory significantly modifies the distribution of the position of the maximum, which is very different from the uniform one, held "a priori". In particular, the "a posteriori" most probable position is at $k = r + 1$.

Example (1.31)

A binary communication channel is composed by a binary input alphabet, say X_0 and X_1, a binary output alphabet, say Y_0 and Y_1, and the matrix of probabilities $P(Y_i/X_i)$, that determine the output symbol Y_i when the input symbol is X_i. The communication problem is a decision problem, that is, to determine which letter among the input alphabet has been transmitted knowing the output letter.

Among the many decision criterions, the soundest one is the one called *Maximum A posteriori Probability* (MAP):

The Bayes' rule allows to write

$$P(X_i/Y_j) = P(Y_j/X_i)\frac{P(X_i)}{P(Y_j)}. \tag{1.32}$$

If we know that a given output has occurred, say Y_0, we say that it has been trasmitted the one among X_0 and X_1 that maximizes (1.32).

For example let us assume that

$$P(Y_0/X_0) = 0.8, \qquad P(Y_1/X_0) = 0.2, \qquad P(Y_0/X_1) = 0.2, \qquad P(Y_1/X_1) = 0.8,$$

(which is called the *Binary Symmetric channel*). Let also assume that input symbols are equally probable. Then also the output symbols are equally probable and from (1.32), maximizing $P(X_i/Y_j)$ means maximizing $P(Y_j/X_i)$.

Assume we receive Y_1. Then, being $P(Y_1/X_1) > P(Y_1/X_0)$ we must decide that X_1 has been transmitted. Similarly, when we receive Y_0 we must decide that X_0 has been transmitted.

The following channel is called perfect channel or noiseless channel.

$$P(Y_0/X_0) = 1, \qquad P(Y_1/X_0) = 0,$$

$$P(Y_0/X_1) = 0, \qquad P(Y_1/X_1) = 1,$$

This means that, when entering X_0, the output is always Y_0 and when entering X_1 the output is always Y_1, so that the decision process is straightforward.

Let consider also the following channel

$$P(Y_0/X_0) = 0.5, \qquad P(Y_1/X_0) = 0.5,$$

$$P(Y_0/X_1) = 0.5, \qquad P(Y_1/X_1) = 0.5,$$

The best decision when receiving Y_0, is left to the reader.

Since error affects all physical measures, and repeating the measure often provides different values, the process toward the interpretation of a measure is exactly a decision process as the one shown above, and shows the importance of the Bayes' rule.

1.7 Statistical independence

Definition. *Two events A and $B \subset S$ are said statistically independent if and only if*

$$P(AB) = P(A)P(B) \tag{1.33}$$

The meaning of statistical independence and the validity of (1.33) is immediate if we observe that if the latter is true, the (1.17) results in:

$$\begin{aligned} P(A/B) &= P(A) \\ P(B/A) &= P(B) \end{aligned}$$

and, in the light of what stated in paragraph 1.4, this means that the probability of e A is is not influenced by the occurrence or non-occurrence of B and vice versa.

One easily checks also that if two events A and B are statistically independent so are their complements $\overline{A}$ and $\overline{B}$. Sometimes the statistical independence can be predicted when probabilistic events correspond to physical events which do not influence each other "physically".

Example (1.34)
Upon the rolling of a dice, check whether the following events

$A = \{$*an even number appears*$\}$
$B = \{$ *number one, or two or three appears*$\}$

are statistically independent.

We have

$$P(A) = 1/2; \qquad P(B) = 1/2; \qquad P(AB) = 1/6;$$

that is

$$P(AB) \neq P(A)P(B)$$

Hence, events A e B are not statistically independent.

Example (1.35)
Upon the rolling of a dice, check whether the following events

$A = \{$*an even number appears*$\}$
$B = \{$*number one, or two, or three, or four appears*$\}$

are statistically independent.

We have

$$P(A) = 1/2; \qquad P(B) = 2/3; \qquad P(AB) = 2/6$$

and

$$P(AB) = P(A)P(B)$$

Hence, events A e B are indeed statistically independent.

The definition of independence is extended to more than two events in the following way:

Definition
events $A_1, A_2, \dots A_n$ are said statistically independent if and only if

$$\begin{aligned} P(A_iA_j) &= P(A_i)P(A_j) \qquad (i, j = 1, 2, \dots n), (i \neq j) \\ P(A_iA_jA_k) &= P(A_i)P(A_j)P(A_k) \qquad (i, j, k = 1, 2, \dots n), (i \neq j \neq k) \\ &\dots\dots\dots \\ &\dots\dots\dots \\ P(A_1A_2\dots A_n) &= P(A_1)P(A_2)\dots P(A_n) \end{aligned} \tag{1.36}$$

The number of relations (1.36) is $2^n n - 1$ and ensure the independence of any number of joint events formed by groups extracted from any $A_1, A_2, \dots A_n$.

1.8 Repeated trials

An experiment such as the n-times flipping of a coin or rolling of a dice presents an outcome that is composed of the outcomes of each single coin or dice. For example, flipping two coins has a space composed of four events: $\{HH\}, \{TH\}, \{HT\}, \{TT\}$. This kind of experiment can be seen as *composed* of subsequent sub-experiments such as the flipping a first coin and the flipping of the second.

Example (1.37)
Rolling two fair *dices check whether X, $X = 1, 2, \dots, 6$, on the first dice and Y, $y = 1, 2, \dots, 6$ on the second are statistically independent events*

The outcome consists on any couple (X, Y) of which there are 36. A specific value $X = x$ on the first dice is the event composed of the six events $\{x, 1\}, \{x, 2\}, \{x, 3\}, \{x, 4\}, \{x, 5\}, \{x, 6\}$. Similar is the event $Y = y$ on the second dice, and their probability is, in a uniform space,

$$P(X = x) = \frac{6}{36}, \qquad P(Y = y) = \frac{6}{36},$$

On the other side the joint event event $\{X = x\}\{Y = y\}$ is composed by only one outcome and

$$P(X = x, Y = y) = \frac{1}{36}.$$

We can easily check that the cited events are statistically independent.♣

The result of the above example can be easily generalized to the rolling of n *fair* dices. In this case it turns out that all events related only to a specific dice are statistically independent from the events related to other dices. Note that in the above experiments we do not distinguish, for example, between the rolling of n dices and n times the rolling of the same dice. When this happens the experiment can be modeled as the repeated drawing of balls from an urn, where each time the drawn ball is reinserted into the urn. More generally we have (the formal proof is omitted)

Theorem: (1.38)
In experiments that can be modeled as repeated drawings from an urn with re-inserction at each draw, events related to different extractions are statistically independent.

Notice that with the urn model we can also perform experiments where the outcomes are not equally probable. Each extraction can be seen as an experiment, and the repeated extractions can be seen as the composition of experiments all described by the same space and probability law. In this case the composed experiment is named *Repeated Trials*.

Note also that the above does not hold for experiments such as the drawing together n cards from a deck of m cards. In fact, this experiment can be lead back to the drawing of balls from an urn, when each time the drawn ball is not reinserted into the urn. In this case, subsequent cards, (urn extractions), are indeed affected by the the preceding extractions.

The theorem above simplifies the evaluation of probabilities of k repeated trials, since the statistical independence let us deal with the space of a single trial, whose cardinality is n, rather then the space of the k trials, whose cardinality is n^k.

Example (1.39)
In rolling three dices, find the probabilities of having faces $\{4, 1, 6\}$ *respectively on the three rollings, and the probability of having such faces whatever their position.*

Since the probability of having a specific face is $1/6$, by the theorem above the first probability is

$$P_1 = \frac{1}{6}\frac{1}{6}\frac{1}{6},$$

while the second is evaluated taking into account that the event is composed of 6 outcomes. Then

$$P_2 = \frac{1}{6}\frac{1}{6}♣$$

Repeated trials with two outcomes, *success* and *failure*, with probability p and q respectively, are called Bernoulli trials. Such are the repeated flipping of a coin or the repeated rolling of a dice whatever face or event we declare as success.

Theorem: (1.40)
The probability of having k successes in n Bernoulli trials with success probability p is given by

$$P(k) = \binom{n}{k} p^k q^{n-k} \qquad (0 \leq k \leq n) \tag{1.41}$$

In fact, each sequence of k successes and $n-k$ failures has probability $p^k q^{n-k}$, whatever the position of the successes. The k successes can be drawn in $\binom{n}{k}$ distinct ways, that represent many disjoint events. The sought probability is then the sum of the probabilities of such sequences with k successes, all of them with probability $p^k q^{n-k}$. ♣

Distribution (1.41), is called *Binomial of order n* probability distribution it represents the generic term of the power expansion of the binomial

$$1 = (p+q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k \, q^{n-k} \tag{1.42}$$

Note that the right hand side represents the probability of the certain event, that is exactly the one.

It is also easy to prove that when k ranges from 0 to n, $P(k)$ initially grows monotonically and then decreases monotonically, reaching its maximum value when $k = k_m$ Integer part of $(n+1)p$. If $(n+1)p$ is an integer we have $P(k_m - 1) = P(k_m)$.

Example (1.43)
A quality control process tests some components out of a factory and components are found defective with a probability ‘a $p = 10^{-2}$. Evaluate the probability that out of 10 components inspected there are

$A = \{$ *only one defective*$\}$
$B = \{$*two defective*$\}$
$C = \{$*at least one defective*$\}$.

The 10 tests can be modeled as Bernoulli trials with success probability $p = \frac{1}{100}$. Then we have

$$P(A) = \binom{10}{1} \left(\frac{1}{100}\right)^1 \left(\frac{99}{100}\right)^9 = 0,0913\ldots$$

$$P(B) = \binom{10}{2} \left(\frac{1}{100}\right)^2 \left(\frac{99}{100}\right)^8 = 0,00415\ldots$$

$$P(C) = \sum_{k=1}^{10} \binom{10}{k} \left(\frac{1}{100}\right)^k \left(\frac{99}{100}\right)^{10-k} = 1 - \binom{10}{0} \left(\frac{1}{100}\right)^0 \left(\frac{99}{100}\right)^{10} = 0,0956\ldots$$

1.8.1 Simulation of repeated trials

Repeated trials, such as the rolling of three dices, or three times the rolling of a single dice, provides as outcome the faces of all the three dices or rollings, corresponding to 6^3 possible outcomes. This means that, in a simulation, the basic experiment must provide at least 6^3 outcomes. The concept of repeated trials, however, allow us to build the basic experiment as the repetition of three basic experiments, representing the independent rolling of each dice.

1.9 Law of large numbers

Dealing with binomial events, such as A and $\bar{A}$, the probability of getting k times event A in n repeated trials is given by (1.41) with $p = P(A)$ and $q = 1 - P(A)$. The same law (1.41) also represents the probability that the frequency F_A, of the occurrence of A over n trials, is k/n. With this law we can prove that

$$\lim_{n\to\infty} P(P(A) - \varepsilon < F_a < P(A) + \varepsilon) = 1 \tag{1.44}$$

for any ε, i.e., as n increases, the probability that the frequency F_A becomes as close as wanted to $P(A)$ tends to one. Unfortunately, probability one does not mean certainty, and this prevents us to get a *physical* definition of probability.

1.10 Problems for solution

P.1.1 Rolling three dices, evaluate the probability of having k equal faces, with $k \in [0; 2; 3]$.

P.1.2 Rolling a dice three times, evaluate the probability of having at least one 6.

P.1.3 Assuming women and men exist in equal number, and assuming that 5% of the men are colour blind and that 0,25% of the women are colour blind, evaluate the probability that a person drawn at random is colour blind. Then evaluate the probability that, having drawn a colour-blind person, this is a male.

P.1.4 Drawn a card from a deck of 52 cards, verify wheter the following events are statistically independent:

a) $A = \{\textit{drawing of a picture card}\};\qquad B = \{\textit{drawing of a hearth card}\}$

b) What if the king of hearths is missing from the deck of cards?
c) What if a card, at random, is missing?

P.1.5 A dice A has four red faces and two white faces. A dice B, vice-versa, has two red faces and four white faces. You flip a coin once, if heads the game continues with dice$A$, otherwise it continues with dice B. a) On rolling the dice, what is the probability that a red face appears on the dice? b) and at the second rolling of the same dice? c) If the first two rollings show a red face, what is the probability that also on the third rolling is red? d) If the first n rollings show a red face, what is the probability that you are using dice A?

P.1.6 An urn contains two white balls and two black. A ball is drawn and replaced with a ball of a different colour. Then a second ball is drawn. Calculate the probability p that the first extracted was white, when the second is white.

P.1.7 The probabilities that three different archers, A, B hit the mark, independently of one another, are respectively 1/6, 1/4 and 1/3. Everyone shoots an arrow. a) Find the probability that only one hits the mark. b) If only one hits the mark, what is the probability he is archer A?

P.1.8 A duel among three people A, B and C is carried out according to the Russian roulette. A six round revolver is loaded with two cartridges. The duelists pass cyclically the weapon, spinning the cylinder every time (so that each duelist has 1/3 probability of being on a loaded chamber) and shooting themselves as long as only one remains alive. Assuming that A is the first, what is the probability that each duelist is the first to die? b) and to win?

P.1.9 From a deck of 52 cards we draw two cards. Find the probabilities of the following events $A = \{\text{the first card is a King; the second figure}\} = \{K_1; F_2\}$, $\quad B = \{\text{at least one figure}\}$, $C = \{\text{the second card is a King}\}$.

Chapter 2

Random Variables

2.1 Spaces with infinite outcomes: Random Variables

Dealing with spaces S with infinite outcomes, to the axioms already exposed in Chapter 1 we must add another one that extends the summation of the probability measure over infinite terms:

Axiom IIIa: *If $A_1, A_2, \ldots, A_n \ldots$ are disjoint events and $A = A_1 + A_2 + \ldots + A_n + \ldots$, then*

$$P(A) = P(A_1) + P(A_2) + \ldots + P(A_n) + \ldots \tag{2.1}$$

An example of such a space is the one whose outcomes are the number N of coin tosses to get a head. This number is not limited, as the head could never appear in n trial, whichever n is. If the coin is fair it can be shown that the probability of the outcomes is

$$P(N = k) = 2^{-k}, \qquad k = 1, 2, 3, \ldots$$

and is such that

$$\sum_k P(N = k) = 1.$$

These spaces are said countables, and can be dealt with the methods exposed in the previous chapter, with some attention to the uniform case, where (1.8) provides 0 for all the events with a finite number of outcomes. This show that we can have events different from the empty set (impossible event) with probability zero, which implies we can have events with probability one that do not coincide with the certain event. In other words

probability zero does not mean impossibility, and probability one does not mean certainty.

If we refer to non-countables spaces, such as is the case when the outcomes is, for example, a point in an interval, or in sets in general geometrical spaces, the problem to assign a probability measure to such spaces is not as simple as it is in countable spaces. In fact, how can we think of a method for assigning a probability to any subset of such a space?

The idea arises to "transform" the space of outcomes into another one where we know how to assign probabilities. In other words, we map outcomes and events (subsets) of S into the space of numbers, integer, real, scalar, or vectorial spaces, which we know well how to manage.

Let at first deal with a real function $f(\alpha)$ defined on the space S of the outcomes. We mean to use a certain law that binds the results $\alpha \in S$ and the set of real numbers R in order to match every α with one and only one value of R; this function is denoted by $f(\alpha)$.

With this function, each set $A \subset S$ corresponds a set $I \subset R$ such that for every $\alpha \in A$ is $f(\alpha) \in I$. In this way the description of an experiment in terms of results α, A and probability events for $P_S(A)$ in S, can be replaced, or rather unified, by the description in terms of real numbers x, sets I and probabilities $P_R(I)$ in R.

A function $f(\alpha)$ which satisfies the above conditions is called "random variable" and is denoted by symbols such as $X(\alpha)$, $Y(\alpha)$, and so on.

2.1.1 Describing a Random Variable

Let X be a Random Variable (RV), as described in the previous paragraph, and x a real number, the probability of the event $\{X \leq x\}$ is a function of the real variable x. Such function is denoted by $F_X(x)$ (or $F(x)$ when there is no doubt on the RV to which it refers) and is called "Cumulative Probability Distribution Function" (CDF) of X. We have thus,

$$F_X(x) = P(X \leq x) \tag{2.2}$$

with the following properties

1. $$F(-\infty) = 0 \qquad F(+\infty) = 1 \tag{2.3}$$

2. is a monotonic non decreasing function of x:
$$F(x_1) \leq F(x_2) \text{ per } x_1 \leq x_2 \tag{2.4}$$

3. is right continuous [1]:
$$F(x^+) = F(x) \tag{2.5}$$

   $F_X(x)$ completely describes RV X; In fact we have, for any x_1, x_2 and x:

4. $$P(x_1 < X \leq x_2) = F(x_2) - F(x_1) \tag{2.6}$$

[1] We denote $F(x^+) = \lim_{\varepsilon \to 0}$ and $F(x^-) = \lim_{\varepsilon \to 0} F(x - \varepsilon)$

5.

$$P(X = x) = F(x) - F(x^-) \tag{2.7}$$

Proofs

- Properties 1 comes from the fact that $\{X \leq -\infty\}$ is the empty set, and $\{X \leq \infty\}$ is the whole space.
- Property (2.4) comes from (1.6) by observing that for $x_1 < x_2$, we have $\{X \leq x_1\} \subset \{X \leq x_2\}$.
- Furthermore, we have $P(X \leq x + \varepsilon) = P(X \leq x) + P(x < X \leq x + \varepsilon)$, and for $\varepsilon \to 0$ the second term tend to zero because the corresponding event becomes the empty set; hence (2.5) [2].
- Property (2.6) comes from the fact that

$$P(x_1 < X \leq x_2) = P(X \leq x_2) - P(X \leq x_1)$$

- Property (2.7) comes from (2.6) by setting $x_1 = x - \varepsilon$, $x_2 = x$ and taking the limit $\varepsilon \to 0$.

We say that two RV X and Y are equal if for every α we have $X(\alpha) = Y(\alpha)$, while we say that they are *equally distributed* if they have the same pdf, ie if $F_X(z) = F_Y(z)$.

We emphasize that given any function $G(x)$ which presents properties corresponding to (2.3) (2.4) and (2.5), we can always build an experiment and define a RV which has $G(x)$ as its CDF. This allows us to treat the distributions without specifying to which experiment it refers.

2.2 Continuous Random Variables

A RV X is said to be of "continuous" type if its CDF $F_X(x)$ is a continuous function in R, together with its first derivative, except at most a countable set of points where the derivative does not exist.

Since for a continuous RV $F_X(x)$ is left continuos, we have from (2.7)

$$P(X = x) = 0.$$

Owing to this, we can usefully introduce the *probability density function* (pdf) of RV X" $f_X(x)$ defined as the derivative of the corresponding CDF:

$$f_X(x) = \frac{dF_X(x)}{dx} \tag{2.8}$$

[2]Notice that if we had set $F(x) = P(X < x)$, this would have been left continuous

The definition is then completed by assigning arbitrary positive values where the derivative does not exist.

From the definition and properties of $F(x)$ we have

$$f(x) \geq 0 \tag{2.9}$$

$$\int_{-\infty}^{\infty} f(x)dx = 1 \tag{2.10}$$

$$P(X \leq x) = F(x) = \int_{-\infty}^{x} f(x)dx \tag{2.11}$$

$$P(x_1 < X \leq x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)dx \tag{2.12}$$

Definition (2.8)shows that

$$f(x) = \lim_{\Delta x \to 0} \frac{P(x < X \leq x + \Delta x)}{\Delta x}. \tag{2.13}$$

This shows that the pdf can be interpreted as the normalized probability that the RV belongs to a small interval around x and, dimensionally, is a density, hence the name.

Example (2.14)

We want to find the CDF and pdf of RV X, defined as the coordinate of a point equally likely to be chosen among others in interval $[a, b]$ of x axis.

As explained in section 2.2 and by (2.2), we immediately have

$$F_X(x) = \begin{cases} \dfrac{x-a}{b-a} & (a \leq x \leq b) \\ \\ 0 & (x < a) \\ 1 & (x > b) \end{cases} \tag{2.15}$$

and from (2.8) or straightly from (2.13)

$$f(x) = \lim_{\Delta x \to 0} \frac{\Delta x}{b-a} \frac{1}{\Delta x}$$

w get

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & (a \leq x \leq b) \\ \\ 0 & \text{elsewehre} \end{cases} \tag{2.16}$$

A RV that satisfies (2.15) and (2.16) is called "uniformly distributed" and the pdf is said "uniform".

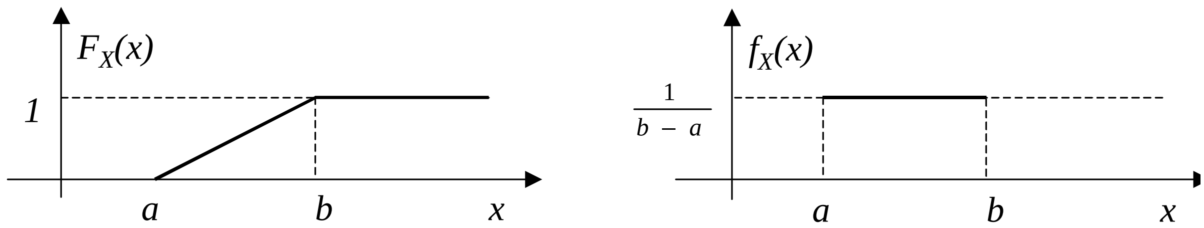


Figure 2.1:

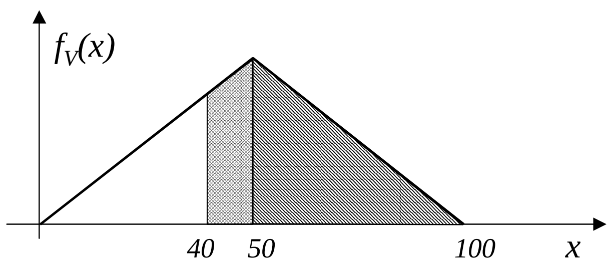


Figure 2.2:

Example (2.17)

Let assume that the pdf of the RV V, defined as the "lifespan" (in years) of human beings, is as depicted in fig. 2.2. We want the probability of the following events:

$A = \{$*death occurs between 40 and 50 years of age*$\}$
$B = \{$*death occurs after 40 years of age* $\}$
$C = \{V \leq t\}$

We observe that $f_V(50) = \frac{1}{50}$, owing to condition (2.10), that assures area 1 for the pdf. From (2.11) and (2.12) we further have

$$P(A) = P(40 < V \leq 50) = \frac{9}{50} \text{ (area ////)}$$

$$P(B) = P(V > 40) = P(A) + \frac{1}{2} = \frac{34}{50} \text{ (area /// and xxx)}$$

$$P(C) = P(V \leq t) = F_V(t) = \begin{cases} 0 & (t \leq 0) \\ \dfrac{t^2}{5000} & (0 < t \leq 50) \\ \dfrac{200t - t^2}{5000} & (50 < t \leq 100) \\ 1 & (t > 100) \end{cases}$$

Example (2.18)

A point P is drawn uniformly on a circumference of radius R and center in the origin of axes.

Find the pdf of RV X, defined as the coordinate of orthogonal projection of P on the horizontal axis.

To find the pdf let us use the (2.13). With reference to Figure 2.3a, $P(x < X \leq x + \Delta x)$ is the probability that P lies in one of two small arcs shown in the figure, each having a length

$$d\ell = \sqrt{dx^2 + dy^2} = dx\sqrt{1 + (\frac{dy}{dx})^2}$$

Being $y = \sqrt{R^2 - x^2}$, by replacing the derivative we get

$$d\ell = \frac{dx}{\sqrt{1 - (\frac{x}{R})^2}}.$$

Then we have:

$$f_X(x) = \lim_{\Delta x \to 0} \frac{1}{\Delta x} \frac{2\Delta l}{2\pi R} = \lim_{\Delta x \to 0} \frac{1}{\Delta x} \frac{1}{2\pi R} \frac{2\Delta x}{\sqrt{1 - (x/R)^2}} = \frac{1}{\pi R} \frac{1}{\sqrt{1 - (x/R)^2}}$$

for ($| x | \leq R$) and zero elsewhere.

The graph of $f_X(x)$ is shown in Figure 2.3b, by which we see that it is more probable to pick the point closer to the extremes rather than to the center.

To obtain the pdf we can also derive the CDF whose expression is found in this case very easily (the task is left to the reader) to give:

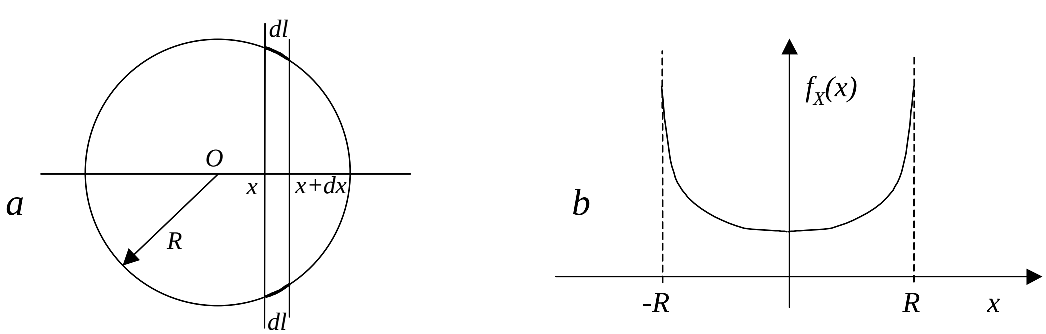


Figure 2.3:

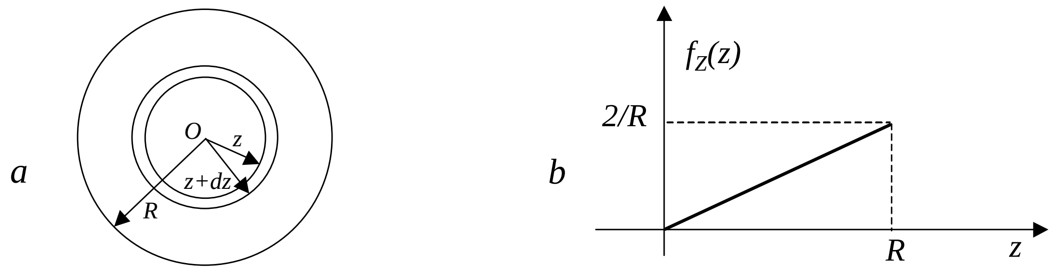


Figure 2.4:

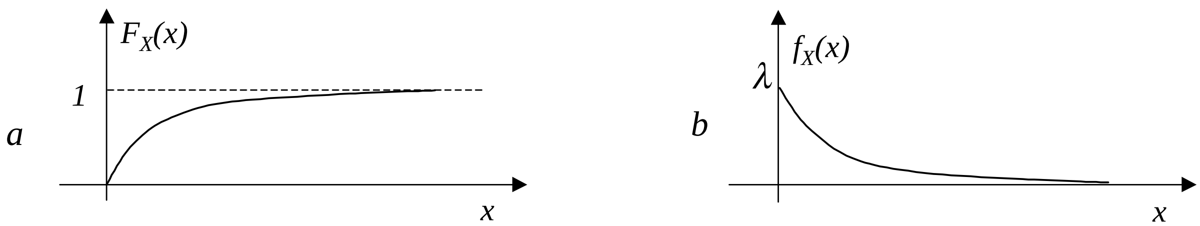


Figure 2.5: CDF and pdf of RV Negative Exponential.

$$P(X \leq x) = F_X(x) = \begin{cases} \dfrac{\pi - arcos\dfrac{x}{R}}{\pi} & (\mid x \mid \leq R) \\ 0 & (x < -R) \\ 1 & (x > R) \end{cases}$$

Example (2.19)

A point P is drawn uniformly in a circle of radius R. Derive the pdf of RV Z, defined as the distance of P from the center O of the circle.

$P(z < Z \leq z + \Delta z)$ is the probability thst P is taken in the annulus shown in figure 2.4a whose area is $2\pi z \Delta z$.

By (2.13) we get

$$f_Z(z) = \lim_{\Delta z \to 0} \frac{2\pi z \Delta z}{\pi R^2} \frac{1}{\Delta z} = \frac{2z}{R^2} \quad (0 \leq z \leq R)$$

The graph of $f_Z(z)$ is shown in Figure 2.4 b. As we can see P is more likely to be selected next to the circumference than at the center.

Example (2.20)

The "Negative Exponential" pdf (figura 2.5) is defined as :

$$F(x) = 1 - e^{-\lambda x} \quad (x \geq 0) \tag{2.21}$$

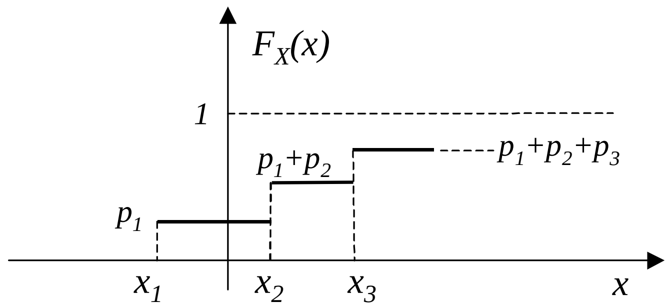


Figure 2.6:

$$f(x) = \lambda e^{-\lambda x} \quad (x \geq 0) \tag{2.22}$$

2.3 Discrete and mixed RV's

A discrete RV X is characterized by a CDF $F_X(x)$ of a staircase type, with discontinuities in a countable set of points $x_i (i = 0, \pm 1, \pm 2 \ldots)$, where it presents jumps of value p_i (figure 2.6). In this case from (2.7) we get

$$P(X = x) = \begin{cases} p_i & x = x_i \\ 0 & x \neq x_i \end{cases} \tag{2.23}$$

(2.23) is called "Probability Distribution of X" and is not to be confused with the CDF $F_X(x)$. We further have

$$p_i \geq 0 \tag{2.24}$$

$$\sum_{i=-\infty}^{\infty} p_i = 1 \tag{2.25}$$

$$F(x) = \sum_{i=-\infty}^{M} p_i \tag{2.26}$$

where M is the maximum i for which $x_i \leq x$.

If in (2.16) the values of x_i are integers, then RV X is said an integer RV. In previous chapters we have already encountered, without explicitly naming them, examples of integer RV's, such as the number that appears on the faces of a dice.

We add here:

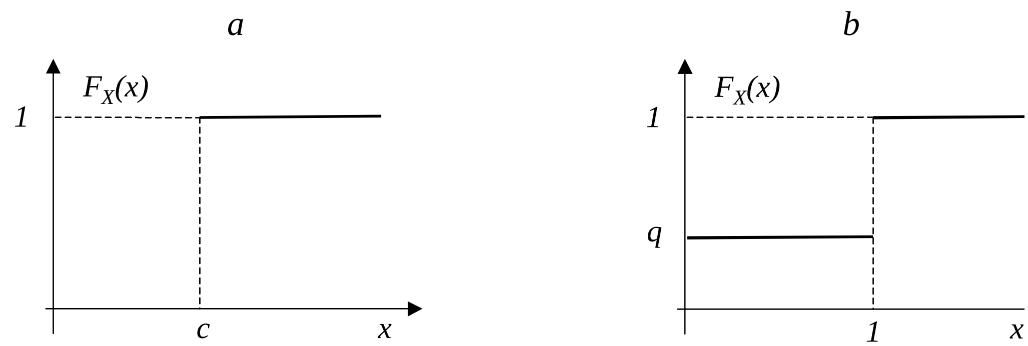


Figure 2.7:

- the distribution of a constant c (figura 2.7a) :

$$P(X = x) = \begin{cases} 1 & \text{for} \quad x = c \\ 0 & \text{elsewhere} \end{cases} \tag{2.27}$$

- the Bernoulli (binary) distribution (figure 2.7b):

$$P(X = x) = \begin{cases} p & \text{for} \quad x = 1 \\ 1 - p = q & \text{for} \quad x = 0 \\ 0 & \text{elsewhere} \end{cases} \tag{2.28}$$

- the uniform distribution

$$P(X = x) = \begin{cases} \frac{1}{n} & \text{for} \quad x = x_i \quad (i = 1, \dots, n) \\ \\ 0 & \text{elsewhere} \end{cases} \tag{2.29}$$

already encountered in the form inside rolls of the dice, draws from urns, etc.

A RV X is said of the "mixed" type if it is not integer and its CDF presents discontinuities (Figure 2.8a).

Such a CDF can always be seen as the sum of a suitable staircase function $F_1(x)$ (figure 2.8b) and a continuous function $F_2(x)$ (figure 2.8c). Examples of such RV will be seen hereinafter.

Here it should be noted that for integer and mixed RV's we can not rigorously define the ddp,which, in some ways, is much more convenient RV's description than the CDF. This difficulty is overcome by resorting to the theory of generalized functions.

The number of Successes in n Bernoulli trials is also a RV whose distribution $P(S_n = k)$ is (1.41):

$$P(S_n = k) = (\begin{smallmatrix} n \\ k \end{smallmatrix}) p^k q^{n-k} \qquad (0 \leq k \leq n) \tag{2.30}$$

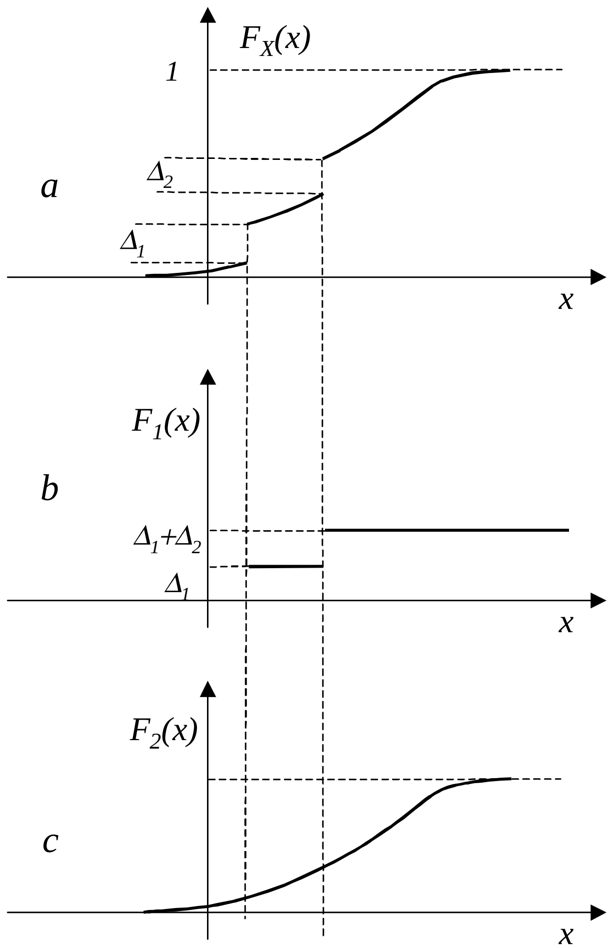


Figure 2.8:

2.4 Moments of a pdf

For the pdf, as for any other function, we can define some parameters that resume some properties of the function. The most used are (if they exist):

1. $k-$th order moments $(k = 1, 2, \ldots)$

$$m_k = \int_{-\infty}^{+\infty} x^k \ f(x)dx \tag{2.31}$$

2. $k-$th order central moments

$$\mu_k = \int_{-\infty}^{+\infty} (x - m_1)^k f(x)dx \tag{2.32}$$

Parameters of the same meaning can be given also for discrete variables in the form:

$$m_k = \sum_{i=-\infty}^{\infty} x_i^k \ p_i \tag{2.33}$$

$$\mu_k \sum_{i=-\infty}^{\infty} (x_i - m_i)^k \ p_i \tag{2.34}$$

In particular, a parameter of paramount importance is the first order moment m_1. This can be interpreted as the coordinate of the center of mass should the pdf represent a mass distribution along the x with density $f(x)$. In other words, it roughly provides the location of the body of the pdf. Similarly, m_2 is a further index of the dispersion of the distribution from the $x = 0$ axis, whereas μ_2 (we have $\mu_1 = 0$) provides an index of the dispersion of the distribution from its proper $x = m_1$ axis. The moments are related to the central moments. For example, by the definition we have

$$\mu_2 = m_2 - m_1^2 \tag{2.35}$$

It can be shown that, if moments of any order exists, then the knowledge of these moments completely the determines the pdf $f_X(x)$ (or the CDF in the discrete RV case).

Example (2.36)

Let us evaluate m_1 and μ_2 for the following RV's

a) Bernoulli RV

b) Binomial RV

a) From (2.33) and (2.34), we get:

$$m_1 = 0 \cdot q + 1 \cdot p = p$$

$$\mu_2 = (0 - p)^2 q + (1 - p)^2 p = pq$$

b)

$$m_1 = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} = \sum_{k=1}^{n} n \binom{n-1}{k-1} p^k q^{n-k} = np \sum_{h=0}^{n-1} \binom{n-1}{h} p^h q^{n-h-1} = np$$

By (2.45) we have

$$\begin{aligned} m_2 &= \sum_{k=0}^{n} k^2 \binom{n}{k} p^k q^{n-k} = np \sum_{k=1}^{n} k \binom{n-1}{k-1} p^{k-1} q^{n-k} \\ &= np \sum_{h=0}^{n-1} (h+1) \binom{n-1}{h} p^h q^{n-h-1} = np(E[X]_{n-1\ prove} + 1) = np[(n-1)p + 1] \end{aligned}$$

$$\mu_2 = m_2 - m_1^2 = npq$$

Example (2.37)
Let us evaluate m_1 and μ_2 for the negative exponential RV. WE have

$$m_1 = \int_0^\infty x\lambda e^{-\lambda x} dx = \frac{1}{\lambda}$$

$$\sigma^2 = m_2 - m_1^2 = \int_0^\infty x^2 \lambda e^{-\lambda x} dx - (\frac{1}{\lambda})^2 = \frac{1}{\lambda^2}$$

Theorem: Law of large numbers (2.38)
If X is an RV whose pdf has first order moment m_1, denoted $X_1, X_2, \ldots, x_n$ the outcomes of the RV in n independent repetitions of the experiment, and $\overline{X}_n$ the arithmetic mean of values of X_i, i.e.,

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

(note that $\overline{X}_n$ is itself a RV) we have:

$$P(\lim_{n\to\infty} \overline{X}_n = m_1) = 1 \tag{2.39}$$

Law (2.39) states that the average performed on a number n of outcomes of n independent trials, tends with probability ‘a 1 to m_1 when n tends to infinity. For this reason, m_1 is also called the "mean value" or" expected value" of RV X and in this sense, it is also denoted by $E[X]$.

The great importance of such law lies in the fact that is provides a relationship between a pure mathematical parameter, m_1, to another one $\overline{X}_n$ drawn from an experiment. Similar laws can also be given for high order moments, if they exists.

The proof of the above law is quite articulated and will not be given here. We rather observe that it can be formulated in another way, that leads to (1.9), tied to the probability p_A of event A. To this purpose, define the binary RV X such that it is $X = 1$ if A occurs and $X = 0$ otherwise. Then, if we perform n trials we have

$$\sum_{i=1}^{n} X_i = n_A$$

being n_A the number of times A occurs. we also observe that

$$m_1(X) = p_A$$

and that

$$\overline{X}_n = \frac{n_A}{n}.$$

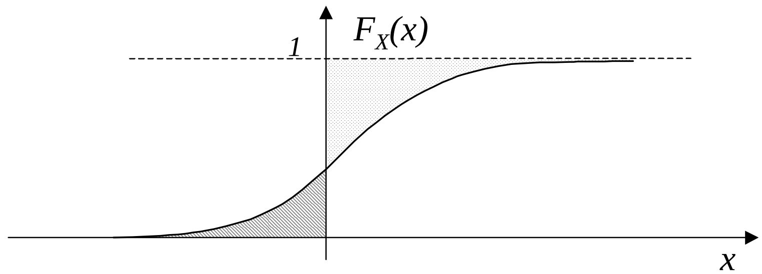


Figure 2.9:

Therefore, the law of large numbers can be written as

$$P(\lim_{n\to\infty} \frac{n_A}{n} = p_A) = 1 \quad \clubsuit \tag{2.40}$$

The above formulation of the law provides the interpretation of probability $P(A)$ as the limit of frequences n_A/n, although this must be considered with probability one.

Other important properties of m_1 are:

1. If $f(x)$ is symmetric around a value of a and m_1 exists, then $m_1 = a$. In fact

$$m_1 = \int_{-\infty}^{+\infty} xf(x)dx = \int_{-\infty}^{+\infty} (y+a)f(y+a)dy = \int_{-\infty}^{+\infty} yf(y+a)dy + a = a$$

This come from (2.10) and from the observation that $f(y+a)$ is an even function ;

2. If m_1 exists, it can be expressed as

$$m_1 = \int_0^{\infty} (1-F(x))dx - \int_{-\infty}^{0} F(x)dx \tag{2.41}$$

that is, nd this as the difference between the dashed and the dotted areas in Figure 2.9.

To prove (2.41) assume at first that $f(x)$ is greater than zero only within $a \le x \le b$. Then m_1 can be written as :

$$m_1 = \int_a^0 xf(x)dx + \int_0^b xf(x)dx$$

Taking the integration by parts and using $F(x)$ e $F(x)-1$ as primitive of $f(x)$ respectively in the first and second integrals we have:

$$m_1 = [xF(x)]_a^0 - \int_a^0 F(z)dx + [x(F(x)-1)]_0^b + \int_0^b (1-F(x))dx$$

Since we have assumed $F(a) = 0$ e $F(b) = 1$

$$m_1 = \int_0^b (1-F(x))dx - \int_a^o F(x)dx$$

and (2.41) comes out by setting $a = -\infty$ and $b = +\infty$;

3. If $F_X(x) = 0$ for $x < 0$, for $\alpha > 0$ the following inequality holds

$$P(X \geq \alpha) \leq \frac{E[X]}{\alpha} \tag{2.42}$$

In fact,

$$E[X] = \int_0^\infty xf(x)dx \geq \int_\alpha^\infty xf(x)dx \geq \alpha \int_\alpha^\infty f(x)dx = \alpha P(X \geq a)$$

hence the thesis. Setting $v = \dfrac{\alpha}{E[X]}$ we get a different expression of (2.42)

$$P(X \geq vE[X]) \leq \frac{1}{v} \tag{2.43}$$

Inequality (ref eq: 3.53) shows how to establish a constraint upon the part of pdf that lies above the mean value ($v > 1$), based on the sole knowledge of the mean value.

$$\mu_1 = \int_{-\infty}^{+\infty} xf(x)dx - m_1 = 0 \tag{2.44}$$

$$\mu_2 = \int_{-\infty}^{+\infty} x^2 f(x)dx - 2m_1 \int_{-\infty}^{+\infty} xf(x)dx + m^2 = m_2 - 2m_1^2 + m_1^2 = m_2 - m_1^2 \tag{2.45}$$

From the definition we see that μ_2 can not be negative; so it must be

$$m_2 \geq m_1^2 \tag{2.46}$$

Central moment μ_2, is also called *variance* of RV X and denoted by σ_X^2, whereas σ_X is called *standard deviation*. The variance represents a measure of the dispersion of $f(x)$ around its average value as shown in the following:

Tchebichev Inequality

when $\mu_2 = \sigma^2$ does exist, we have

$$P(| X - m_1 | \geq v\sigma) < \frac{1}{v^2} \tag{2.47}$$

In fact:

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - m_1)^2 f(x)dx \geq \int_{|x-m_1| \geq v\sigma} (x - m_1)^2 f(x)dx \geq$$

$$\geq v^2\sigma^2 \int_{|x-m_1| \geq v\sigma} f(x)dx \geq v^2\sigma^2 P(| X - m_1 | \geq v\sigma)$$

hence the thesis. By setting $v\sigma = \varepsilon$ we get alternatively

$$P(m_1 - \varepsilon < X < m_1 + \varepsilon) \;\geq\; 1 - \frac{\sigma^2}{\varepsilon^2} \tag{2.48}$$

$$P(|X - m_1| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} \tag{2.49}$$

from which we see that if σ^2 is small, there is a high probability that X belongs to a short interval around m_1. From (2.48) we also see that when $\sigma^2 = 0$, then

$$P(X = m_1) = 1$$

that is, X provides the same constant value for *almost all* the outcomes of the experiment.

Example (2.50)

Let us apply Tchebichev inequality to bound the probability that the frequency of HEADS in flipping a fair coin n times exceeds $0.5 \pm \varepsilon$.

The frequency of HEADS in n trials is H/n where H is the RV number of HEADS in n trials. This has a Binomial distribution with average $n/2$ and $\sigma^2(H) = n/4$. Therefore,

$$m_1(H/n) = \frac{1}{2}$$

$$\sigma^2(H/n) = \frac{1}{4n}$$

Tchebichev inequality says

$$P(|H/n - m_1| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

and substituting

$$P(|H/n - 0.5| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2}$$

we have

$$\begin{array}{llll}
\varepsilon = 0.1, & n = 10, & P & \leq 2.5(???) \\
\varepsilon = 0.1, & n = 100, & P & \leq 0.25 \\
\varepsilon = 0.1, & n = 1000, & P & \leq 0.025 \\
\varepsilon = 0.1, & n = 10000, & P & \leq 0.0025
\end{array}$$

We also see that

$$\lim_{n\to\infty} P(|H/n - 0.5| \geq \varepsilon) = 0, \qquad \forall \varepsilon > 0$$

that provides one kind of demonstration of the law of large numbers.♣

Other interesting parameters are

- absolute moments

$$m'_k = \int_{-\infty}^{+\infty} \mid x^k \mid f(x)dx \tag{2.51}$$

- generalized moments

$$\mu_k^{(a)} = \int_{-\infty}^{+\infty} (x-a)^k f(x)dx \tag{2.52}$$

- the mode x_m, defined as the most probable value, i.e., the value at which the pdf is maximum;
- the median m, defined as the value exceeded (or not exceeded) with probability 0.5. In symmetrical pdf the median coincides with the mean m_1.

Example (Persistence of bad luck) (2.53)

RV X represents the measure of the misfortune experienced in a certain circumstance or trial (waiting time, financial loss, etc..). Denoted by X_0 the misfortune I experienced and $X_1, X_2, \dots, X_n, \dots$ the misfortune experienced by others after me in subsequent and independent trials. Denoted by N the number of the first trial in which one is more unfortunate than me (i.e., N is the smallest value of n for which $X_n > X_0$), we want to determine the distribution and average value of RV N.

$P(N = n)$ is the probability that among $n+1$ trials the most unfortunate trials lies in n and the second most unfortunate trials lies in 0 (me). The latter event can happen in $(n-1)!$ different ways among the $(n+1)!$ possible outcomes. Therefore we have

$$P(N = n) = \frac{(n-1)!}{(n+1)!} = \frac{1}{(n+1)n} \qquad (n = 1, 2, \dots)$$

and

$$E[N] = \sum_{n=1}^{\infty} nP(N = n) = \sum_{n=1}^{\infty} \frac{1}{n+1} = \infty$$

My bad luck is then without limit!

Example (2.54)

A gambling is said fair *if the average RV V gain (that is negative if one actually loses) is zero. Chech wether the bet on the outcome of a roulette number is fair game.*

Denoted by C the amount of the bet, the average gain is

$$V = \begin{cases} 35C & \text{with probability } \dfrac{1}{37} \\ \\ -C & \text{with probability } \dfrac{36}{37} \end{cases}$$

$$E[V] = 35C\frac{1}{37} - C\frac{36}{37} = -\frac{1}{37}C$$

The game is not fair, as the bank gets an average gain equal to $C/37$.

2.5 Conditional Distributions and Densities

Since distributions and densities represent probabilities, we can easily extend to them the conditional definition.

Let M be an event of space S where RV X is defined. We define CDF of X conditional to M (provided that $P(M) \neq 0$) the function:

$$F_X(x/M) = P(X \ \leq \ x/M) \tag{2.55}$$

and similarly for the density a

$$f_X(x/M) = \frac{dF_X(x/M)}{dx} = \lim_{\Delta x \to 0} \frac{P(x < X \leq x + \Delta x/M)}{\Delta x} \tag{2.56}$$

It is easy to check that the above defined functions have all the properties of the CDF and pdf.

In particular, we can define the conditional average

$$E[X/M] = \int_{-\infty}^{+\infty} x f_X(x/M) dx \tag{2.57}$$

and, in the same way, any other conditional moments.

Also the total probability theorem can be extended to distributions, densities and expected values:

$$F_X(x) = \sum_i F_X(x/M_i) P(M_i), \tag{2.58}$$

$$f_X(x) = \sum_i f_X(x/M_i) P(M_i), \tag{2.59}$$

$$E[X] = \sum_i E[X/M_i] P(M_i), \tag{2.60}$$

Relevant cases are those when even event M is described in terms of RV X (especially dealing with RVs representing the life of people, the service life of a certain machine or components, the duration of any phenomenon in general).

Example (2.61)
With reference to Example 2.17, evaluate and draw

a) the pdf of the lifespan V of people that has reached forty years of age;

b) the pdf of the RV Z "remaining lifespan" of people that has reached forty years of age.

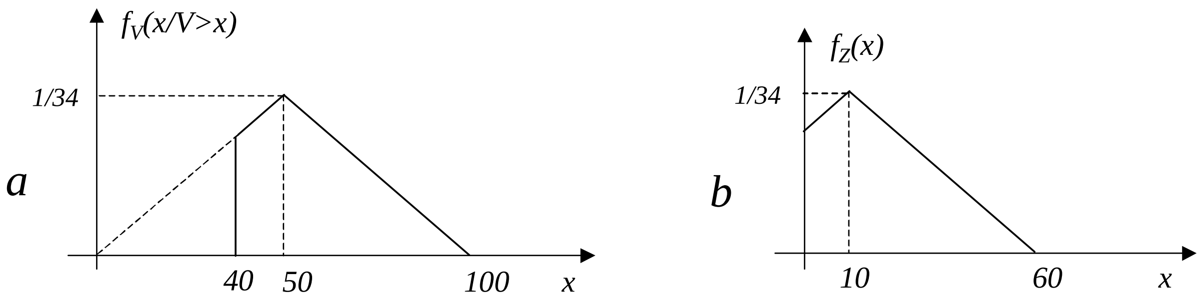


Figure 2.10:

a) The pdf is the one of V conditioned to event C, and from (2.56) and (1.17) we get

$$\begin{aligned} f_V(t/V>40) &= \lim_{\Delta t\to 0}\frac{P(t<V\le t+\Delta t, V>40)}{P(V>40)}\frac{1}{\Delta t} = \lim_{\Delta t\to 0}\frac{P(t<V\le t+\Delta t)}{P(V>40)}\frac{1}{\Delta t} = \\ &= \frac{f_V(t)}{1-F_V(40)} = \frac{50}{34}f_V(t), \qquad \text{for } \ 40<t\le 100 \text{ and } 0 \text{ elsewhere.} \end{aligned}$$

The corresponding graph is the one represented by the solid line in fig. 2.10a; this coincides with a portion of the originl pdf, re-normalized as to still provide area one.

b) We have

$$\begin{aligned} f_Z(t) &= \lim_{\Delta t\to 0}\frac{P(t<Z\le t+\Delta t)}{\Delta t} = \lim_{\Delta t\to 0}\frac{P(t<V-40\le t+\Delta t/V>40)}{\Delta t} = \\ &= \lim_{\Delta t\to 0}\frac{P(t+40<V\le t+40+\Delta t/V>40)}{\Delta t} = f_V(t+40/V>40). \end{aligned}$$

From the result in a) we get

$$f_Z(t) = \frac{f_V(t+40)}{1-F_V(40)}, \qquad (0<t\le 60),$$

that is, $f_Z(t)$ coincides with the pdf in a) after a translation on the left by 40 years (fig. 2.10b).

2.6 Events conditional to the values of a RV

We define probability of an event A conditional to the value x assumed by a RV X, assuming that $f_X(x)\neq 0$, the limit

$$P(A/X=x) = \lim_{\Delta x\to 0} P(A/x<X\le x+\Delta x). \tag{2.62}$$

From Bayes formula (1.27) we get:

$$P(A/X = x) = \lim_{\Delta x \to 0} \frac{P(x < X \leq x + \Delta x/A)P(A)}{P(x < X \leq x + \Delta x)},$$

and multiplying by Δx above and below, and taking the limit, from (2.13) and (2.56) we have finally

$$P(A/X = x) = \frac{f_X(x/A)P(A)}{f_X(x)}. \tag{2.63}$$

In this way, definition (1.17) is extended also to the case where $P(M) = P(X = x) = 0$, provided that $M = \{X = x\} \neq \phi$. From (2.63) we have then

$$\int_{-\infty}^{+\infty} f_X(x/A)P(A)dx = \int_{-\infty}^{+\infty} P(A/X = x)f_X(x)dx,$$

and, by observing that $\int_{-\infty}^{+\infty} f_X(x/A)dx = 1$, we have

$$P(A) = \int_{-\infty}^{+\infty} P(A/X = x)f_X(x)dx. \tag{2.64}$$

Furthermore from (2.63), and using (2.64), we finally obtain

$$f_X(x/A) = \frac{P(A/X = x)f_X(x)}{P(A)} = \frac{P(A/X = x)f_X(x)}{\int_{-\infty}^{+\infty} P(A/X = x)f_X(x)dx}. \tag{2.65}$$

Relations (2.64) e (2.65) represent respectively the theorem of Total Probability (1.20) and the Bayes theorem (1.28) extended to the continuous case.

Example (2.66)

The distance of two points A and B on a circumference of length L is a RV X with uniform pdf. Another point C is chosen in a uniform way and independently on the circumference. Find the probability of event $\{C \in \widehat{AB}\}$. If we assume $X = x$, we have

$$P(C \in \widehat{AB}|X = x) = \frac{x}{L}, \qquad 0 \leq x \leq L.$$

From (2.64), and being $f_X(x) = \frac{1}{L}$, $(0 \leq x \leq L)$, we get

$$P(C \in \widehat{AB}) = \int_0^L P(C \in \widehat{AB}|X = x)f_X(x)dx = \int_0^L \frac{x}{L^2}\,dx = \frac{1}{2}$$

Since the points are taken in an uniform way, the result can also be found also observing that, once A is taken, the two permutations of A and B are equally likely.

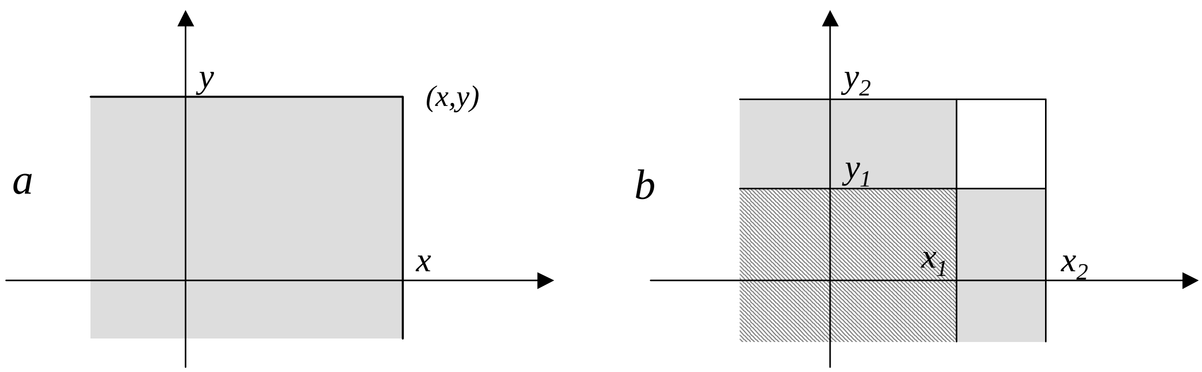


Figure 2.11:

2.7 Vectorial RVs

We extend here the concepts and definitions of a scalar RV climb to a pair of RV's, being immediate extension to the case of more than two RV's

Consider two RV $X(\alpha)$ and $Y(\alpha)$ defined in the same result space S (e.g. the coordinates of a point in the plane, the height and the weight of a person, etc..). By means of this pair of functions a correspondence arises between each event $A \subset S$ and a set D_{xy} of the Cartesian plane, such that for every $\alpha \in A$ the point with coordinates $X(\alpha)$ and $Y(\alpha)$ belongs to D_{xy}, a joint event in S is thus represented by a domain D_{xy} in the Cartesian plane.

The probability of the joint events (Figure 2.11a)

$$\{X \leq x, Y \leq y\} = \{X \leq x\}\{Y \leq y\}$$

Is a function of the pair of real variables x and y. Such a function, denoted by $F_{X,Y}(X, y)$, called the "joint CDF RVs' X and Y.

We have then

$$F_{XY}(x, y) = P(X \leq x, Y \leq y) \tag{2.67}$$

From the definition we can easily verify the following relations:

$$F(x, \infty) = F_X(x); \qquad F(\infty, y) = F_Y(y) \tag{2.68}$$

$$F(\infty, \infty) = 1 \tag{2.69}$$

$$F(x, -\infty) = 0; \qquad F(-\infty, y) = 0 \tag{2.70}$$

$$P(x_1 < X \leq x_2, y_1 < Y \leq y_2) = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) \tag{2.71}$$

The latter property can be easily deduced from (2.67) observing Figure2.11b.

relations (2.68), (2.69), (2.70) represent sufficient conditions to let a function of two variables represent a joint CDF. Assuming now that $F_{XY}(x, y)$ has the derivatives that are needed, the joint pdf of RVs X and Y is

$$f_{XY}(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} \tag{2.72}$$

From the properties previously described we also have

$$f(x, y) \geq 0 \tag{2.73}$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1 \tag{2.74}$$

Furthermore, from the definition of the joint derivative we have

$$f(x, y) = \lim_{\Delta x, \Delta y \to 0} \frac{P(x < X < x + \Delta x, y < Y < y + \Delta y)}{\Delta x \Delta y} \tag{2.75}$$

Relation (2.75) will be often used as starting point to derive the pdf seeked in some problems.

Denoted by $\{(X, Y)\}$ the event of all results α where $X(\alpha)$ and $Y(\alpha)$ belong to domain D, it can be written as a union or intersection of elementary events of the type

$$\{x < X \leq x + \Delta x,\ y < Y \leq y + \Delta y\}$$

and, therefore, we have

$$P((X, Y) \in D) = \int \int_D f(x, y) dx dy \tag{2.76}$$

where the integral is extended over the domain D. It also follows

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy; \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx \tag{2.77}$$

The concept of joint pdf can then be extended to the case of discrete and mixed RVs by proceeding as in Chapter 2.

When dealing with multiple RVs, the distributions and the densities of a single RV are called *marginal* to emphasize the difference with joint CDFs and pdf's.

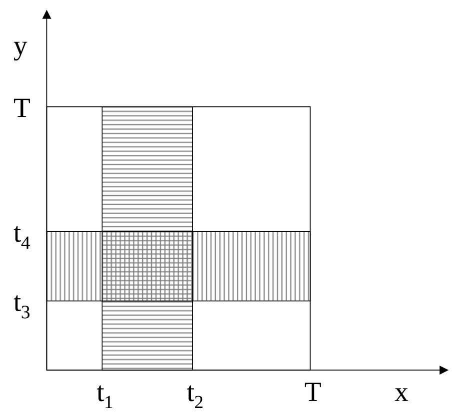


Figure 2.12:

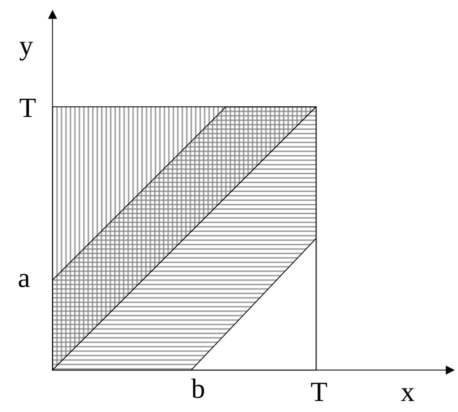


Figure 2.13:

Example (2.78)
A train and a passenger arrive at a station at times that are represented by RVs X and Y whose joint pdf is uniform in $0 \le x \le T$ and $0 \le y \le T$. The trains stops for a time span of a minutes, whereas the passenger stop at a station for a time span of b minutes, after which, if the train has arrived, it takes the train; otherwise he leaves the station. We want to evaluate the probability of the following events:

$A = \{t_1 < X < t_2\}$,
$B = \{t_3 < Y < t_4\}$,
$C = \{t_1 < X < t_2\} \cdot \{t_3 < Y < t_4\}$,
$D = \{X < Y\}$,
$E = \{$*the passenger is able to take the train*$\}$.

The outcomes of the experiment are represented by points (X, Y) in the square in fig. 2.12. In the same figure we have represented events A (horizontal hatching), B (vertical hatching), and $AB = C$ (cross hatching). Since with a uniform pdf the probability of a domain is the ratio of the domain area to the universe area, we have

$$P(A) = \frac{(t_2 - t_1)T}{T^2} = \frac{t_2 - t_1}{T}; \qquad P(B) = \frac{t_4 - t_3}{T}; \qquad P(C) = \frac{(t_2 - t_1)(t_4 - t_3)}{T^2}.$$

Events D and E are represented in fig. 2.13 with, respectively, vertical and horizontal hatching. The event $X = Y$ is represented by the points on the bisector of the quadrant, that splits the square into two equal areas. Therefore $P(D) = 1/2$.

Finally, we have $E = \{Y - a < X < Y\} + \{X - b < Y < X\}$, so that

$$P(E) = \frac{a + b}{T} - \frac{a^2 + b^2}{2T^2}.$$

Example Buffon's needle (2.79)
A needle of length ℓ is thrown on a ruled paper, where lines are spaced by $d > \ell$. Evaluate the probability that the needle crosses a line, assuming that the position of the needle's center and its angle with the line are two RVs jointly uniform.

We can determine the position of the needle's center by its distance to the closest line, i.e., we can assume that the center always lies within a strip whose length is $d/2$ (fig. 1.10). In the same way, the angle ϕ changes from zero to $\pi/2$. The space of the experiment is, then, the rectangle shown in fig. 2.15.

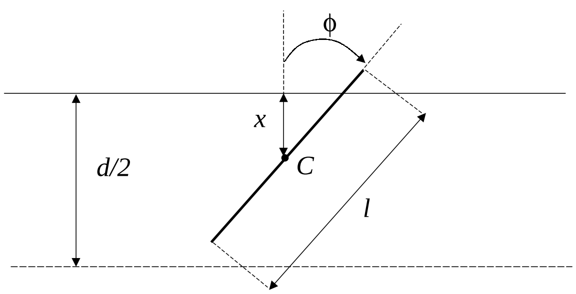


Figure 2.14:

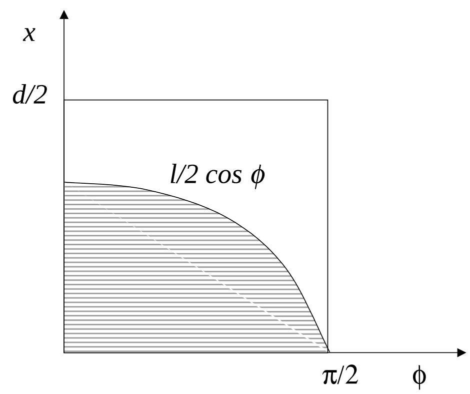


Figure 2.15:

The event is verified when $x < \frac{\ell}{2}\cos\phi$, represented in the figure by the hatched area. The probability is then

$$P = \frac{2\ell}{\pi d}.$$

Example Ordered statistic. (2.80)
Two points A and B are uniformly selected in the interval $[0, T]$. Find the pdf and the average of RV X, defined as the coordinate of the point (either A or B) that happens to be the closest to the origin.

Denoting by H and K the RVs coordinates of A and B, their joint pdf is uniform in the square of side T. In order to have $\{x < X \leq x + \Delta x\}$ either point must lie in $[x, x + \Delta x]$, and the other one must lie in $[x + \Delta x, T]$. The area of the event in the space (h, k) is $2\Delta x(T - x - \Delta x)$, and we have

$$f(x) = \lim_{\Delta x \to 0} \frac{P(x < X \leq x + \Delta x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{1}{\Delta x} \frac{2\Delta x(T - x - \Delta x)}{T^2} = \frac{2(T - x)}{T^2}, \qquad 0 \leq x \leq .$$

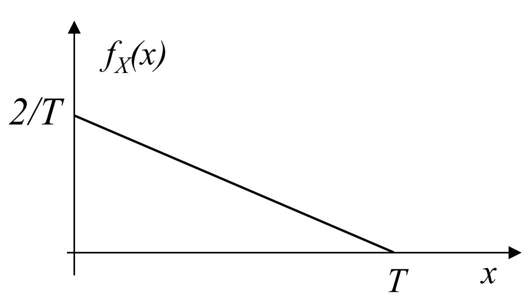


Figure 2.16:

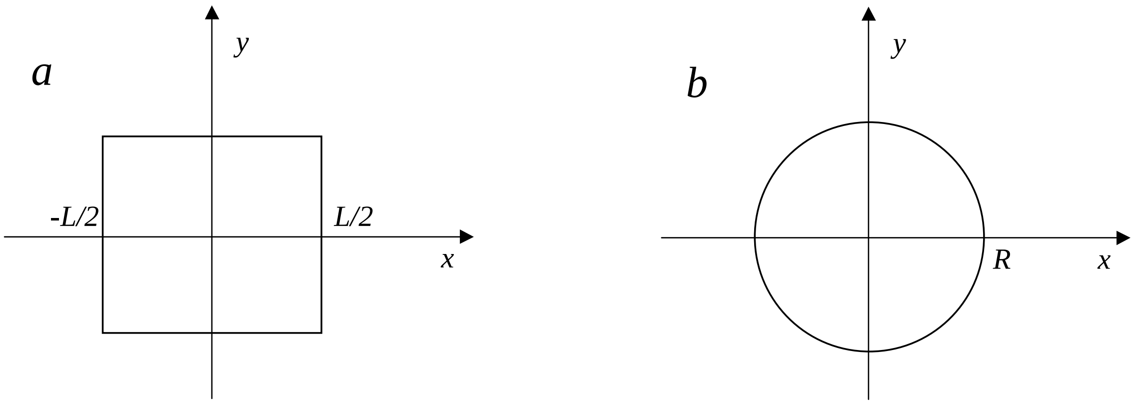


Figure 2.17:

The density is shown in figure 2.16, and average value is found to be $E[X] = T/3$.

It is left to the reader to find the pdf and the average value of Y, the coordinate of the point (either A or B) that happens to be the farthest to the origin. The reader is also encouraged to generalize the problem to n uniform points on the same interval.

Example (2.81)
Find the joint and marginal pdf of RV's X and Y Cartesian coordinates of a point Q chosen uniformly in a

a) *square of side L and centered at the origin (Figure 2.17a)*

b) *circle of radius R and center at the origin (Figure 2.17b)*

To find the joint density we use (2.75). In this expression the probability at the numerator the probability Q lies into the rectangle of coordinates x, $x + \Delta x$, y, $y + \Delta y$, but since Q is picked uniformly, this probability takes value $\frac{\Delta x \Delta y}{S}$, S being the area of the domain, regardless of the location of the small rectangle. Therefore, we obtain

$$f(x, y) = \begin{cases} \frac{1}{S} & \text{for}(x, y) \in S \\ \\ 0 & \text{elsewhere} \end{cases} \tag{2.82}$$

Such a pdf is still called Uniform in S and the value of the constant $1/S$ depends only from the area of the domain and not by its shape.

About the marginal pdf we have
a)

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y)dy = \int_{-\frac{L}{2}}^{\frac{L}{2}} \frac{1}{L^2} dy = \frac{1}{L}; \qquad \left(-\frac{L}{2} < x < \frac{L}{2}\right)$$

and similarly

$$f_Y(y) = \frac{1}{L}; \qquad \left(-\frac{L}{2} < y < \frac{L}{2}\right)$$

In this case, the marginal pdf are uniform.
b)

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)dy = \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \frac{1}{\pi R^2}dy = \frac{2}{\pi R^2}\sqrt{R^2-x^2}; \qquad (\mid x \mid < R)$$

Here, the marginal pdf's are no longer uniform. In fact, the shape of the domain of point (X, Y) influences the result.

Example (2.83)
Find the joint and marginal pdf's of RV's Z and Θ, polar coordinates of the point A in the previous example.

Again from (2.75), noting that the area of the elementary surface whose corners have coordinates z, $z + dz$, θ, $\theta + d\theta$ (see Figure 2.18a) amounts to $dz \times zd\theta$, we get

$$f_{Z\Theta}(z,\theta) = \frac{z}{S} \qquad ((z,\theta) \in S)$$

We then have

b)

$$f(z,\theta) = \frac{z}{\pi R^2}; \qquad (0 \le z < R)(-\pi < \theta \le \pi)$$

$$f_Z(z) = \int_{-\pi}^{\pi} f(z,\theta)d\theta = \frac{2}{R^2}z \qquad (0 \le z < R)$$

$$f_\Theta(\theta) = \int_0^R f(z,\theta)dz = \frac{1}{2\pi} \qquad (-\pi < \theta \le \pi)$$

a)

$$f(z,\theta) = \frac{z}{L^2} \qquad (\mid z\cos\theta \mid < \frac{L}{2}, \mid z\sin\theta \mid < \frac{L}{2})$$

In evaluating $f_z(z)$ we must note that for $z < \dfrac{L}{2}$ the integration interval in θ is $-[\pi;\pi]$, which yields

$$f_Z(z) = \frac{2\pi}{L^2}z;$$

on the other side, for $\dfrac{L}{2} \le z < L\dfrac{\sqrt{2}}{2}$ the integration interval is shown by the dash sectors in Figure 2.18b; this yields

$$f_Z(z) = \frac{4}{L^2}(arcsen\frac{L}{2z} - arccos\frac{L}{2z})z$$

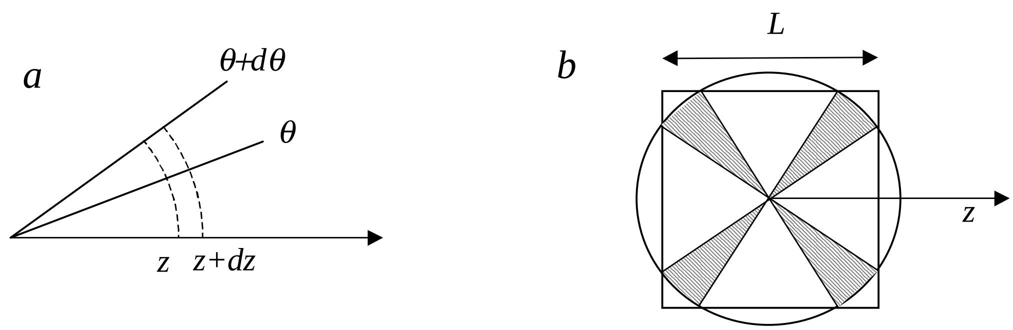


Figure 2.18:

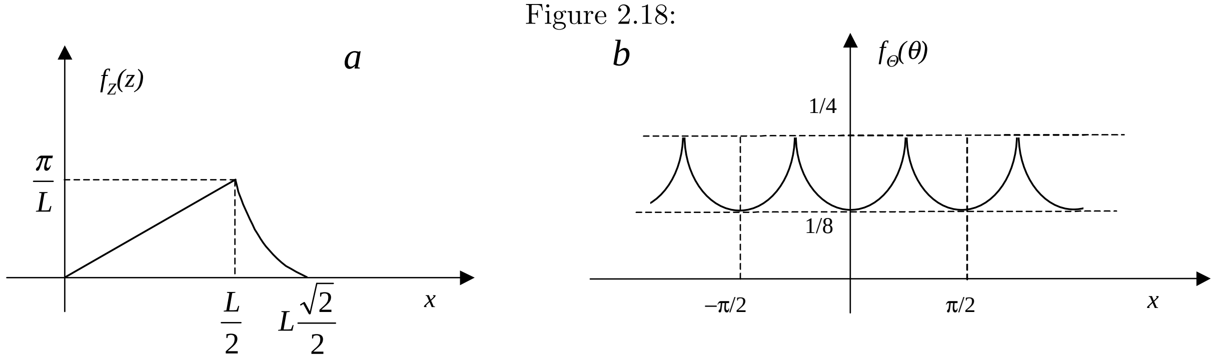


Figure 2.19:

The complete graph is shown in Figure 2.19a.

Then, about RV θ we have

$$f_\Theta(\theta) = \int_0^{|L/2\cos\theta|} f(z,\theta)dz = \frac{1}{8\cos^2\theta} \qquad (-\pi < \theta \leq -\frac{3}{4}\pi)(-\frac{\pi}{4} < \theta \leq \frac{\pi}{4})(\frac{3}{4}\pi < \theta \leq \pi)$$

$$f_\Theta(\theta) = \int_0^{|L/2\sin\theta|} f(z,\theta)dz) = \frac{1}{8\sin^2\theta} \qquad (-\frac{3}{4}\pi < \theta \leq -\frac{\pi}{4})(\frac{\pi}{4} < \theta \leq \frac{3}{4}\pi)$$

The graph is shown in Figure 2.19b.

2.8 Conditional pdf's

The extension to the case of two or more RV's of definitions and theorems of paragraph 2.7 are obtained immediately. Here, we limit ourselves to show the new cases of

• pdf of RV Y conditioned by the value assumed by another RV X

$$f_Y(y/X=x) = \frac{f_{XY}(x,y)}{f_X(x)} \tag{2.84}$$

- Total Probability Theorem

$$f_Y(y) = \int_{-\infty}^{+\infty} f_Y(y/X = x) f_X(x) dx \tag{2.85}$$

- Bayes'Ttheorem

$$f_Y(y/X = x) = \frac{f_X(x/Y = y) f_Y(y)}{f_X(x)} \tag{2.86}$$

- conditional mean

$$E[Y/X = x] = \int_{-\infty}^{+\infty} y f_Y(y/X = x) dy \tag{2.87}$$

- Total Probability Theorem with respect to the mean

$$E[Y] = \int_{-\infty}^{+\infty} E[Y/x] f_X(x) dx \tag{2.88}$$

The demonstration of the relations shown above, their validity in the average values as well as the extension to the case of more RV's, is obtained in the usual way.

Example (2.89)
Two trains leave the central station at given times, represented by points A and B respectively, on a circumference (clock) of length L. A passenger, arriving at the station at a uniform time instant C in the day, wants take the first train that is leaving. Find the pdf of the waiting time W, i.e., the time it has to wait before the first train starts, and its average.

Let $\widehat{AB}$ and $\widehat{BA}$ be the two oriented arcs determined by the points. We use the total probability theorem in the form (2.59)

$$f_W(w) = f_W(w|C \in \widehat{AB}) P(C \in \widehat{AB}) + f_W(w|C \in \widehat{BA}) P(C \in \widehat{BA}),$$

where, owing to the uniform arrival of the passenger, we have

$$P(C \in \widehat{AB}) = \frac{L_{\widehat{AB}}}{L}, \qquad P(C \in \widehat{BA}) = \frac{L_{\widehat{BA}}}{L}.$$

For the same reason the conditional waiting time of the passenger is uniform:

$$f_W(w|C \in \widehat{AB}) = \frac{1}{L_{\widehat{AB}}}, \qquad 0 \leq w \leq L_{\widehat{AB}},$$

and similarly, for the other conditional density. We finally have

$$f_W(w) = f_W(w|C \in \widehat{AB}) P(C \in \widehat{AB}) + f_W(w|C \in \widehat{BA}) P(C \in \widehat{BA}),$$

which can be written as

$$f_W(w) = \frac{1}{L} \left(\mathrm{rect}_{L_{\widehat{AB}}}(w) + \mathrm{rect}_{L_{\widehat{BA}}}(w) \right),$$

where

$$\mathrm{rect}_T(w) = 1, \qquad 0 \le w \le T.$$

The average can be derived again by the Total probability theorem (2.60)

$$E[W] = E[W|C \in \widehat{AB}]P(C \in \widehat{AB}) + E[W|C \in \widehat{BA}]P(C \in \widehat{BA}) =$$

$$= \frac{L_{\widehat{AB}}}{2}\frac{L_{\widehat{AB}}}{L} + \frac{L_{\widehat{BA}}}{2}\frac{L_{\widehat{BA}}}{L} = \frac{L^2_{\widehat{AB}} + L^2_{\widehat{BA}}}{2L}.$$

We draw the attention on the square law above, that can be rather unexpected and with unexpected consequences. For example, the minimum is attained when $L_{\widehat{AB}} = L_{\widehat{BA}} = L/2$, with $E[W] = L/4$, and the maximum for $L_{\widehat{AB}} = 0$ and $L_{\widehat{BA}} = L$, with $E[W] = L/2$.

Example (2.90)

Waiting time paradox. The departure times, A and B, of the trains of Example 2.89 are taken randomly and in an uniform way on L. Again, find the pdf and the average of the waiting time to the first departure W.

We proceed as in Example 2.89 where, now, $L_{\widehat{AB}}$ and $L_{\widehat{BA}}$ are random variables. Setting $X = L_{\widehat{AB}}$, X presents a uniform in $[0, L]$, whereas $L_{\widehat{BA}} = L - X$. Conditioning on the value of $X = x$, let us use the results of Example 2.89, namely

$$f_W(w|X = x) = \frac{1}{L}\left(\mathrm{rect}_x(w) + \mathrm{rect}_{L-x}(w)\right),$$

$$E[W|X = x] = \frac{x^2 + (L - x)^2}{2L}$$

Now, by the total probability theorem in (2.85), we have

$$f_W(w) = \int_0^\infty f_W(w|X = x)f_X(x)dx = \frac{1}{L}\int_0^L \left(\mathrm{rect}_x(w) + \mathrm{rect}_{L-x}(w)\right)\frac{1}{L}dx =$$

$$= \frac{1}{L^2}\int_w^L dx + \frac{1}{L^2}\int_0^{L-w} dx = 2\frac{L - w}{L^2}, \qquad 0 \le w \le L.$$

In a similar way we have o apply the Total Probability Theorem (2.88) referred to the average,

$$E[W] = \int E[W/X = x]f_X(x)dx = \int_0^L \frac{x^2 + (L - x)^2}{2L^2}dx = \frac{L}{3}$$

We encourage the reader to verify that the above results are exactly those of Example 2.81.

If $L = 24$ hours, then we have $E[W] = 8$ hours. Note that should the trains leave at equal distances (12 hours) we would have $E[W] = 6$, which is, in fact, the intuitive results at first would one expect.

Example (2.91)

A point of coordinate X is uniformly taken within interval $[0; L]$ of x axis; Another point of coordinate Y is uniformly taken within interval $[X; L]$. Find the joint pdf of X, Y, the marginal pdf of Y and the probability P that the three segments of length X, $Y - X$, and $L - X$ can form a triangle.

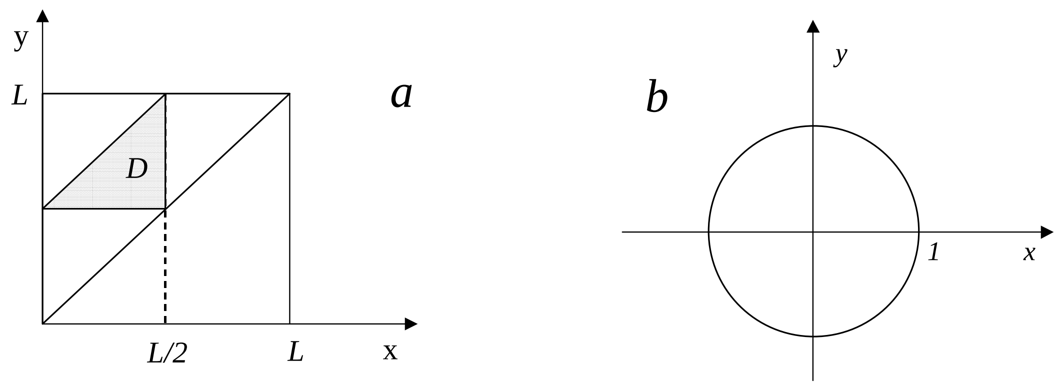


Figure 2.20:

We have

$$f_X(x) = \frac{1}{L}; \qquad (0 < x < L)$$

$$f_Y(y/X = x) = \frac{1}{L - x}; \qquad (x < y < L)$$

Using (2.84) we get

$$f_{XY}(x, y) = \frac{1}{L(L - x)}; \qquad (0 < x < y < L)$$

and from (2.85), by observing that the expression under integration is zero for $x > y$, we have

$$f_Y(y) = \int_0^y \frac{1}{L(L - x)} dx = \frac{1}{L} ln \frac{L}{L - y}; \qquad (0 < y < L)$$

The domain D where X and Y are such as to allow the construction of the triangle, each segment shorter than the sum of the other two, is shown in Figure 2.20a,) and we thus have

$$p = \int_0^{\frac{L}{2}} dx \int_{\frac{L}{2}}^{x+\frac{L}{2}} \frac{1}{L(L - x)} dy = ln2 - \frac{1}{2} = 0,1931\ldots$$

Example (2.92)

Find the joint and marginal pdf's of the coordinates X and Y of a point P uniformly chosen on a circumference of unit radius and center at the origin 0. (Figura 2.20b)

The marginal pdf of X has been already evaluated in Example 2.18 as

$$f_X(x) = \frac{1/\pi}{\sqrt{1 - x^2}}; \qquad (| x | \leq 1)$$

and by symmetry $f_Y(y) = f_X(y)$.

The joint density is degenerate because is different from zero only on the circumference $x^2 + y^2 = 1$. You can, however, make use (2.84) and the impulse function, to write

$$f_Y(y/X=x) = \frac{1}{2}\,\delta(y-\sqrt{1-x^2}) + \frac{1}{2}\,\delta(y+\sqrt{1-x^2})$$

In fact, for $X = x$, Y can take only the two values $\pm\sqrt{1-x^2}$ withe same probability.

From (2.84) we obtain

$$f_{XY}(x,y) = \frac{1}{2}(\delta(y-\sqrt{1-x^2}) + \delta(y+\sqrt{1-x^2}))\frac{1/\pi}{\sqrt{1-x^2}}$$

or, by simmetry

$$f_{XY}(x,y) = \frac{1}{2}(\delta(x-\sqrt{1-y^2}) + \delta(x+\sqrt{1-y^2}))\frac{1/\pi}{\sqrt{1-y^2}}.$$

2.9 Statistically independent RV's

Two RV X and Y are said to be statistically independent if events $\{X \leq x\}$ e $\{Y \leq y\}$ are statistically independent for each x and y.

It follows then that two random RV are independent if one of the following relations holds

$$\begin{aligned}
&F_{XY}(x,y) = F_X(x)F_Y(y), && \forall(x,y),\\
&f_{XY}(x,y) = f_X(x)f_Y(y), && \forall(x,y),\\
&f_X(x/Y=y) = f_X(x), && \forall(x,y),\\
&f_Y(y/X=x) = f_Y(y), && \forall(x,y).
\end{aligned}$$

Similarly, from (1.36) when we have more than two RV's.

Example (2.93)
Check which pairs of RV's treated in previous examples are statistically independent.

In Example 2.81 X and Y are independent in case a) but not in the case b), whereas in Example 2.83 Z and θ are independent in b) but not in a). As we see, the statistical dependence may be linked to the choice of the variables used to describe a phenomenon. Example 2.91 X and Y are not (obviously) independent. Example 2.92 X and Y are tied together by a deterministic relation.

Example (2.94)
Given two RV's X and Y independent and exponentially distributed with the same average $1/\lambda$, find:

a) the probability of the event $\{Y > \alpha X\}$ with α real positive;

b) the pdf $f_Y(y/Y > \alpha Y)$.

a) We could use the (2.76), being D the domain in which $y > \alpha x$, and given the independence we have

$$f_{XY}(x,y) = f_X(x) f_Y(y) = \lambda^2 e^{-\lambda(x+y)} \quad (x, y > 0)$$

More immediately we can use the Total Probability Theorem

$$\begin{aligned} P(Y > \alpha X) &= \int_0^\infty P(Y > \alpha X / X = x) f_X(x) dx = \int_0^\infty e^{-\lambda \alpha x} \, \lambda e^{-\lambda x} dx = \\ &= \frac{1}{\alpha+1} \int_0^\infty \lambda(\alpha+1) e^{-\lambda(\alpha+1)x} dx = \frac{1}{(\alpha+1)} \end{aligned}$$

b) From the definition of conditional pdf, and from the result of point a) we get:

$$\begin{aligned} f_Y(y/Y > \alpha X) &= \lim_{\Delta y \to 0} \frac{1}{\Delta y} \frac{P(y < Y \le y + \Delta y, Y > \alpha X)}{P(Y > \alpha X)} = \\ &= \lim_{\Delta y \to 0} \frac{1}{\Delta y} \frac{P(y < Y \le y + \Delta y, X < y/\alpha)}{P(Y > \alpha X)} = \\ &= \frac{\displaystyle\int_0^{y/\alpha} f_{XY}(x,y) dx}{P(Y > \alpha X)} = \frac{\lambda e^{-\lambda y} \displaystyle\int_0^{y/\alpha} \lambda e^{-\lambda x} dx}{1/(\alpha+1)} = \\ &= (\alpha+1) \lambda e^{-\lambda y} (1 - e^{-\lambda y/\alpha}) \quad (y > 0) \end{aligned}$$

The same result could be easily found by using the Bayes' Theorem in the following way:

$$f_Y(y/Y > \alpha X) = P(Y > \alpha X / Y = y) \, \frac{f_Y(y)}{P(Y > \alpha X)} \tag{2.95}$$

Note that the limit expressions of the derived pdf for $\alpha \to 0$ and $\alpha \to \infty$ are respectively $\lambda e^{-\lambda y}$ e $\lambda^2 y e^{-\lambda y}$ (Erlang-2)♣.

An important example that involves independent random variables is the one related to Repeated Trials, where on each trial i an RV, X_i is defined. We saw in Section 1.8 that events related to only one trial are statistically independent from others. This means also that a specific X_i is independent from others.

2.10 Joint Moments of two RV's

Given two RV's X and Y the joint moments of order h and k are defined as

$$m_{hk} = \int \int x^h y^k f_{xy}(x,y) dx dy$$

and the central moments of order h and k

$$\mu_{hk} = \int \int (x - m_x)^h (y - m_y)^k f_{xy}(x, y) dx dy.$$

Note that the marginal moments occur when one of the two indices h or k is zero; they coincide with the moments of the same order of the single RV that is, for example:

$$m_{0k} = \int \int y^k f_{xy}(x, y) dx dy = \int y^k f_y(y) dy$$

$$m_{h0} = \int x^h f_X(x) dx$$

and also

$$\begin{aligned} m_{10} &= m_X \\ m_{01} &= m_Y \\ \mu_{20} &= \sigma_X^2 \\ \mu_{02} &= \sigma_Y^2 \end{aligned}$$

The mixed second-order central moment mu_{11}, said also Covariance of V.C. X and Y, is of f particular interest. It is linked to m_{11} by the following relation

$$\mu_{11} = m_{11} - m_{10} m_{01} \tag{2.96}$$

From the definitions we see that if X and Y are statistically independent the integral splits into the product of two separate integrals, and we have

$$m_{hk} = m_{h0} \cdot m_{0k}$$

$$\mu_{hk} = \mu_{h0} \cdot \mu_{0k}$$

Specifically, for independent RVs (2.96) shows that

$$\mu_{11} = 0 \tag{2.97}$$

Note, however, that zero covariance does not imply independence among the variables. We only sat that RVs with zero covariance are *uncorrelated*. Also, the following inequalities are valid.

$$m_{11}^2 \leq m_{20}\ m_{02} \tag{2.98}$$

$$\mu_{11}^2 \leq \mu_{20}\ \mu_{02} \tag{2.99}$$

2.11 Problems for solution

Chapter 2

P.2.1 Given the function $f(x) = \dfrac{C}{\alpha^2 + x^2}$, determine the relationship between C e α in order to make $f(x)$ a pdf. (Cauchy).

P.2.2 A point P uniformly chosen in a square of Side L centered at the origin and the x-axis. Find the pdf of RV X, coordinate of the orthogonal projection of P on the horizontal axis.

P.2.3 A point P uniformly chosen in a circle of radius R centered at the origin and the x-axis. Find the pdf of RV X, coordinate of the orthogonal projection of P on the horizontal axis.

P.2.4 Find the first order moment of pdf $f(x) = \lambda^2 x e^{-\lambda x}, x \geq 0$, and 0 elsewhere.

P.2.5 Find the first order moment of the integer distributions

1. $P(X = k) = (1-p)^{k-1}p, \quad k \geq 1;$
2. $P(X = k) = (1-p)^k p, k \geq \quad 0.$

P.2.6 A coin of radius r is flipped on a grid paper where $d > 2r$ is the size of the grid squares. If the center of the coin can be taken in a uniform way on the paper, find the probability the coin covers a square vertex.

P.2.7 Two points are chosen uniformly and independently in a segment of length L. Find the probability that the three obtained segments can be arranged in a triangle.

P.2.8 Referring to Example 2.80, find the joint pdf of (X, Y) and the marginal pdf of $W = Y - X$.

P.2.9 Two points are chosen uniformly and independently in a circle of radius R. Find the pdf of RV X distance to the center of the point closest to the center.

P.2.10 Take a number X from one to six, throw three dices. You win C if X appears once, $2C$ if X appears twice, $3C$ if it appears three times, and you lose C if X does not appear. Check whether this is a fair game.

P.2.11 Assume the RV X, lifespan of a component, is uniform in $[0; L]$. We know that the component age is z; find the pdf of its lifespan. Find the pdf of Y, remaining lifespan.

P.2.12 Repeat the previous exercise assuming that the pdf of X is negative exponential. Find the fair amount a a customer of age z must pay to get a capital C if he dies before the year.

P.2.13 Check whether functions of x and y below can represent joint pdfs' and if so check whether X and Y are statistically independent.

1. $f(x, y) = 4xy \qquad (0 \leq x \leq 1;\ 0 \leq y \leq 1),$
2. $f(x, y) = 8xy \qquad (0 \leq x \leq y;\ 0 \leq y \leq 1),$
3. $f(x, y) = 4x^2 y \qquad (0 \leq x \leq 1;\ 0 \leq y \leq 1)$

P.2.14 A person in phone booth makes a phone call whose duration is represented by RV X, with negative exponential pdf with mean value $1/\mu$. A second person comes after a time Y, RV negative exponentially with average $1/\lambda$, independent of X. Find the pdf of RV W, the time the latter has to wait to the end of the call.

P.2.15 Given two independent RVs' X, and Y, find the probability of the event $\{Y \leq X\}$ when

1. f_X, f_Y are uniform within intervals respectively $[-1; 3], [0; 4]$;

2. f_X, f_Y with the same pdf (you do not need to know the pdf).

What about event $\{Y \leq X/2\}$?

P.2.16 Find the pdf of RV $Z = \min(X, Y)$, where X and Y are two independent negative exponential RVs' with parameters λ and μ respectively. (Hint: observe that $\min(X, Y) > z$ if $x > z$ and $Y > z$. Also, we may take the condition $Y = y$...)

P.2.17 Take interval $[0, X]$, where X is a RV Erlang-2. Then take a point P uniformly within the preceding interval. Find the pdf of Y, length of $\overline{0P}$.

P.2.18 n points are uniformly taken within $[0; T]$. Find the probability that k out of n point lie within an interval $[0; X]$ where RV X is uniform in $[0; T]$.

P.2.19 Two RVs' X and Y are independent and uniformly distributed in $[0; 1]$. Find $f_X(x|X > Y)$, $f_{XY}(x, y|X > Y)$ and $P(X > 2Y|X > Y)$.

P.2.20 Chosen four points A, B, C and D in a uniform way and independently on a circumference, find the probability that chords AB and CD cross each other.

Chapter 3

Functions of RV's

If we have a RV $X(\alpha)$ and a single valued function $y = g(x)$ defined in the domain of X, then if at each outcom α we take the value $g(X(\alpha) = Y(\alpha)$, Y is a random variable and we say it is a function of RV X.

3.1 CDF of a function of RV

Let I_y be the set on the real axis such that $g(x) \leq y$ if and only if $x \in I_y$, then we have

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \in I_y) \tag{3.1}$$

Since I_y represents a probabilistic event, $P(X \in I_y)$ can be expressed as the sum and difference of values of $F_X(x)$.

Example (3.2)

Let RV X be uniform within $-C$ and C. Find the cdf of RV $Y = \dfrac{1}{X^2}$.

We have

$$g(x) = \frac{1}{x^2}$$

The set I_y, $y > 0$ is composed of intervals (Figure 3.1a)

$$x \leq x_1 = -\frac{1}{\sqrt{y}}, \qquad x \geq x_2 = \frac{1}{\sqrt{y}},$$

whereas for $y < 0$ it is $I_y = \phi$.

From (3.1) we get

$$F_Y(y) = P(X \leq -\frac{1}{\sqrt{y}}) + P(X \geq \frac{1}{\sqrt{y}}) = F_X(-\frac{1}{\sqrt{y}}) + 1 - F_X(\frac{1}{\sqrt{y}}), \qquad y > 0$$

$$F_Y(y) = 0, \qquad y \leq 0$$

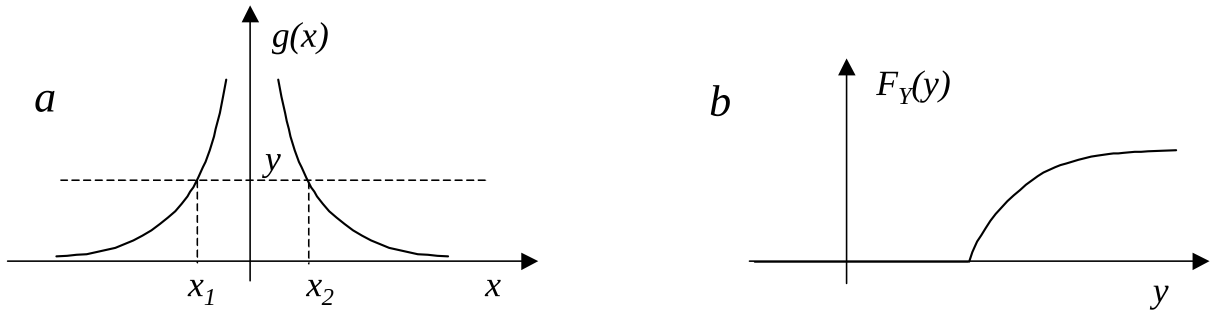


Figure 3.1:

Starting from

$$F_X(x) = \begin{cases} 1 & (x > C) \\ \dfrac{x}{2C} + \dfrac{1}{2} & (\mid x \mid \leq C) \\ 0 & (x < -C) \end{cases}$$

and observing that condition $\mid x \mid \leq \ C$ corriesponds to $y \geq \dfrac{1}{C^2}$, we finally get (Figure 3.1b)

$$F_Y(y) = \begin{cases} 0 & (y \leq 0) \\ 0 & (0 < y < \dfrac{1}{C^2}) \\ 1 - \dfrac{1}{C\sqrt{y}} & (y \geq \dfrac{1}{C^2}) \clubsuit \end{cases}$$

Easy to represent is the case where Y is a discrete variable.

Example (3.3)
Let RV X have the cfd shown in Figure 3.2a. Find the cdf of RV $Y = g(X)$ where

$$g(x) = \begin{cases} 1 & (x > 0), \\ -1 & (x \leq 0). \end{cases}$$

Function $g(x)$ is a stepwise function (Figure 3.2b); then Y is a discrete RV that can only take values ± 1. We have:

$$P(Y = -1) = P(X \leq 0) = F_X(0),$$

$$P(Y = 1) = P(X > 0) = 1 - F_X(0).$$

The resulting $F_Y(y)$ is shown in Figure 3.2c.$\clubsuit$

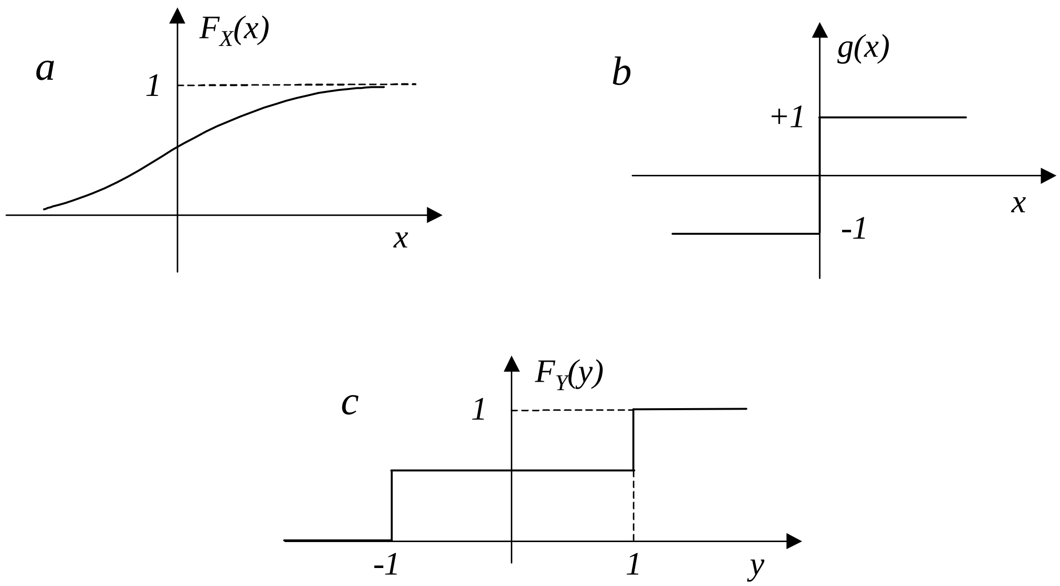


Figure 3.2:

A very important case is the one where $g(x)$ coincides with the CDF $F_X(x)$, strictly increasing, of a continuous RV X. In this case also $Y = g(X)$ is a continuous RV and we have $Y \leq y$ only when $X \leq x$, and $P(Y \leq y) = P(X \leq x)$. Since is by definition $y = F_X(x) = P(X \leq x)$, we also have $P(Y \leq y) = y$, $0 \leq y \leq 1$. Hence we have the

Theorem: (3.4)
The RV Y obtained as the function $Y = F_X(X)$, is an uniform RV in $[0; 1]$.

Note that this theorem show us how to transform a uniform variable Y in $[0; 1]$ into a variable X whose $F_X(x)$ is strictly increasing. We must take $x = g(y)$ where G is the inverse function of $y = F_X(x)$.

3.2 PDF of a function of RV

The pdf of a function Y of RV X can be obtained by taking the derivative of its cdf, in turn attained as shown above. However, we can get the pdf of Y, when it exists, directly from the pdf of X.

Theorem: (3.5)
Let X be a continuous RV and $g(x)$ a continuous function whose derivative exists everywhere, never being equal to zero other that in a countable set of points (it can not have horizontal segments of finite length).

The conditions assumed make RV Y continuous. Denoted by $x_1(y), x_2(y), \ldots, x_n(y), \ldots$ the real roots of equation

$$y = g(x), \tag{3.6}$$

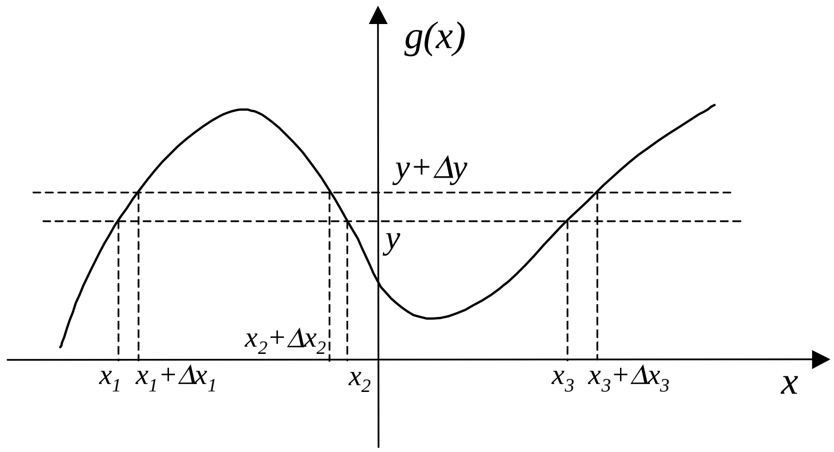


Figure 3.3:

The pdf of $Y = g(X)$*,is:*

$$f_Y(y) = \begin{cases} \sum_n \dfrac{f_X(x_n)}{\mid g'(x_n) \mid} \\ \\ 0 \quad \textit{if no root exists.} \end{cases} \tag{3.7}$$

Proof

Le consider the $g(x)$ in Figure 3.3 that shows the roots of (3.6) x_1, x_2, x_3, and let $x_1+\Delta x_1$, $x_2-\Delta x_2$, $x_3 + \Delta x_3$ be the roots in correspondence of $y + \Delta y$. We the have

$$P(y < Y \leq y + \Delta y) = P(x_1 < X \leq x_1 + \Delta x_1) + P(x_2 - \Delta x_2 < X \leq x_2) + \dots$$
$$\frac{P(y < Y \leq y + \Delta y)}{\Delta y} = \frac{P(x_1 < X \leq x_1 + \Delta x_1)}{\Delta x_1} / \frac{\Delta y}{\Delta x_1} + \frac{P(x_2 - \Delta x_2 < X \leq x_2)}{\Delta x_2} / \frac{\Delta y}{\Delta x_2} + \dots$$

Taking the limit $\Delta y_1 \to 0$, by our hypotheses we also have Δx_1, $\Delta x_2, \ldots \to 0$. The (3.7) turns out thanks to the (2.13) and to the following:

$$\lim_{\Delta x_n \to 0} \frac{\Delta y}{\Delta x_n} =\mid g'(x_n). \mid .$$

Example (3.8)

Find the pdf of RV $1/X$ *where* X *is a RV uniform within* $[a; b]$.

The solution is explained in Figures 3.4a, b e c. Equation $y = g(x)$ $(x > 0)$ has only one solution:

$$x_1 = \frac{1}{y}, \qquad y > 0.$$

We also have

$$g'(x) = -\frac{1}{x^2},$$

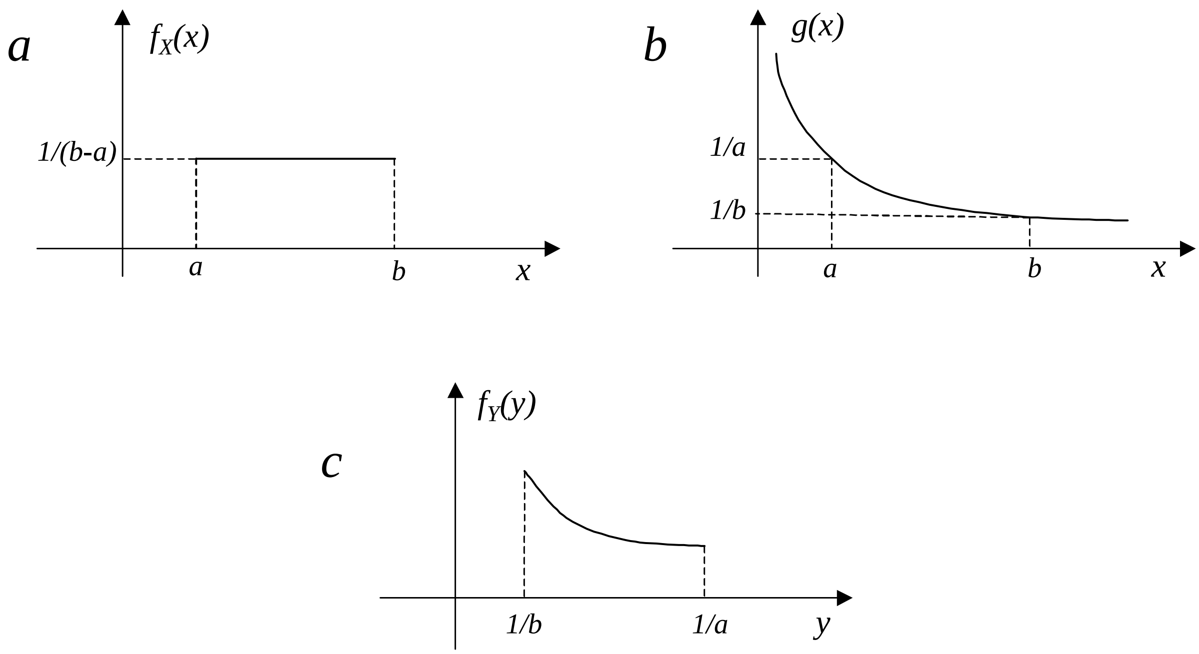


Figure 3.4:

$$g'(x_1) = -y^2.$$

From (3.7) we finally get

$$f_Y(y) = \begin{cases} \dfrac{1}{(b-a)y^2} & (\dfrac{1}{b} \leq y \leq \dfrac{1}{a}) \\ 0 & \text{elsewhere.} \clubsuit \end{cases}$$

Example (3.9)
Find the pdf of RV $1/\sqrt{X}$ where X is a RV negative exponential.

We have

$$f_X(x) = \lambda e^{-\lambda x} \quad (x \geq 0,)$$

$$g(x) = \sqrt{x} \quad (x \geq 0),$$

$$g'(x) = \frac{1}{2\sqrt{x}} \quad (x > 0,)$$

For $y > 0$ the only solution of (3.6)(Figure 3.5) is $x_1 = y^2$, and in correspondence we have

$$g'(x_1) = \frac{1}{2y},$$

$$f(x_1) = \lambda e^{-\lambda y^2}.$$

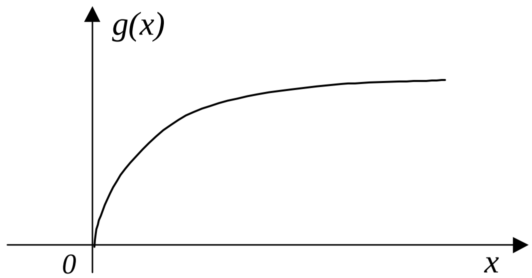


Figure 3.5:

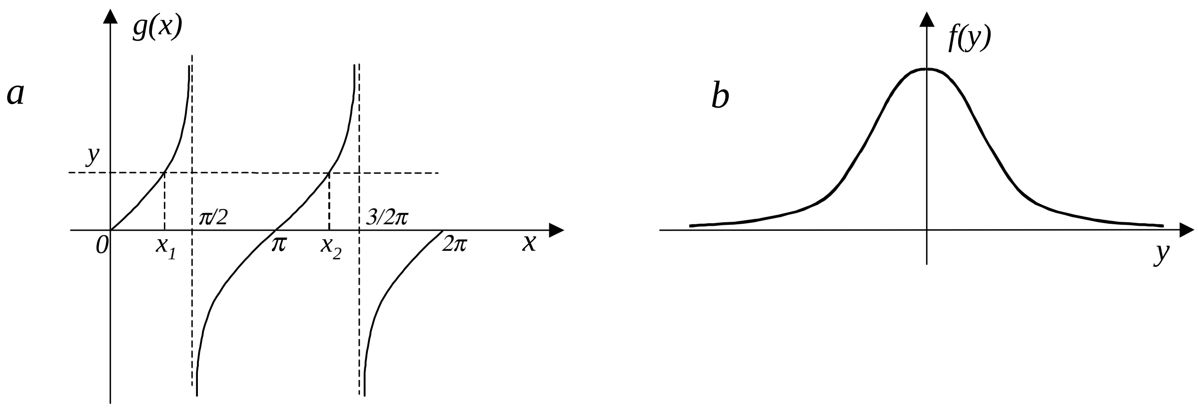


Figure 3.6:

From (3.7) we finally get

$$f_Y(y) = 2\lambda y e^{-\lambda y^2} \quad (y > 0)$$

which is called ”Rayleigh”.

Example (3.10)
Find the pdf of RV $\tan X$ *where* X *is uniform in* $[0; 2\pi]$.

The solutions of (3.6) $x_n = \arctan y$ are infinite in number and differ by π (Figure 3.6a). We have

$$g'(x) = 1 + \tan^2 x,$$

$$g'(x_n) = 1 + y^2, \forall\ n).$$

$$f_X(x) = \begin{cases} \dfrac{1}{2\pi} & (0 \le x < 2\pi) \\ \\ 0 & \text{altrove.} \end{cases}$$

We have only two roots x_n in $[0; 2\pi]$; then we have only two ters in the summation above, yielding

$$f_Y(y) = \frac{1}{\pi}\ \frac{1}{1+y^2}. \tag{3.11}$$

Note that we get the same result if X is uniform in $[0; \pi]$. Such a pdf (Figure 3.6b) is called "Cauchy". Cauchy pdf has the property of remaining unchanged for an inverse trasformation, i.e., for $Z = 1/Y$ we get

$$f_Z(x) = f_Y(x).$$

Also note that Cauchy pdf do not present any moment of any order.♣

Here we provides a simple anlternative proof of theorem (3.4). Function $g(x) = F_X(x)$ is strictly increasing and in (3.7) we have only one term corresponding to the lone solution x_1:

$$f_Y(y) = \frac{f_X(x_1)}{F'_X(x_1)} = \frac{F'_X(x_1)}{F'_X(x_1)} = 1. \quad (0 \leq y < 1).$$

Therefore, RV Y is uniform within $[0; 1]$.

3.3 The sum of two continuous RV's

Given the two continuous RV's X e Y, whose joint pdf is known, we want to find the pdf of heir sum

$$Z = X + Y \tag{3.12}$$

To this purpose, we note that

$$f_Z(z/X = x) = f_Y(z - x/X = x) \tag{3.13}$$

From the total probability theorem we have

$$f_Z(z) = \int f_Z(z/X = x) f_X(x) dx = \int f_Y(z - x/X = x) f_X(x) dx \tag{3.14}$$

which provides the final formula

$$f_Z(z) = \int f_{XY}(x, z - x) dx \tag{3.15}$$

Symmetrically we have

$$f_Z(z) = \int f_{XY}(z - y, y) dy \tag{3.16}$$

If X and Y are statistically independent the two above become

$$f_Z(z) = \int f_X(x) f_Y(z - x) dx \tag{3.17}$$

$$f_Z(z) = \int f_X(z - y) f_Y(y) dy \tag{3.18}$$

The operation in (3.17) and (3.18) are known as the convolution of pdf's. In fact, the convolution of functions $f(x)$ and $g(y)$ (need not to be pdf's) is defined as

$$f(z) * g(z) = \int f(x)g(z-x)dx = \int f(z-x)g(x)dx$$

Example (3.19)

Find the pdf of RV $Z = X + Y$ where X and Y are independent RVs with the same pdf, namely

a) $f(x) = \dfrac{1}{a} \qquad (0 < x < a)$

b) $f(x) = \lambda e^{-\lambda x} \qquad (x > 0)$

a) The integrating function in (3.17) is different from zero when both the following conditions apply:

$$\begin{cases} 0 < x < a \\ 0 < z - x < a \end{cases}$$

or

$$\begin{cases} 0 < x < a \\ z - a < x < z \end{cases}$$

Such conditions depend on z and, therefore, we must distinguish the following cases:

- for $z \le 0 \qquad f_Z(z) = 0$
- for $0 \le z < a$ condition $0 < x < z$ holds, and therefore we have
$$f_Z(z) = \frac{1}{a^2}\int_0^z dz = \frac{z}{a^2};$$
- for $a < z \le 2a$ condition $z - a < x < a$ holds, and therefore we have
$$f_Z(z) = \frac{1}{a^2}\int_{z-a}^{a} dx = \frac{2-z}{a^2};$$
- for $z > 2a \qquad f_Z(z) = 0;$

The seeked pdf is shown in Figure 3.7.

b) The integrating function in (3.17) is different from zero when $\begin{cases} x > 0 \\ z - x > 0 \end{cases}$ that is $\begin{cases} x > 0 \\ x < z \end{cases}$ and, therefore, we have

$$f_Z(z) = \int_0^z \lambda e^{-\lambda x}\lambda e^{-\lambda(z-x)}dx = \lambda^2 z e^{-\lambda z} \quad (z > 0)$$

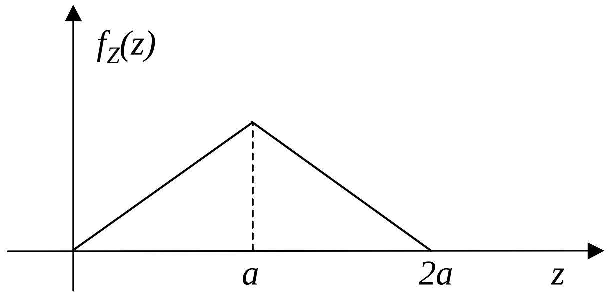


Figure 3.7:

The pdf we get is called Erlang-2. This is a pdf out of family

$$E_k(x) = f_k(x) = \frac{(\lambda x)^{k-1}}{(k-1)!}\lambda e^{-\lambda x}, \qquad x \geq 0 \tag{3.20}$$

called the Erlang family. Actually the Erlang-1 coincides with the negative exponential. Indeed we have

$$E_m = E_1^{*m} \tag{3.21}$$

$$E_m * E_n = E_{m+n} \tag{3.22}$$

Property (3.22) can be stated saying that the Erlang family is closed with respect to the convolution.

3.4 The sum of two integer RV's

Much as in the previous case we can write the distribution of the sum $Z = X + Y$, with X and Y integer RV's as

$$P(Z = k) = \sum_j P(X = j, Y = k - j) = \sum_j P(X = k - j, Y = j), \tag{3.23}$$

which, again, becomes the convolution if the two RV's are independent.

Example (3.24)

Find the distribution of RV $Z = X + Y$ where X and Y are Binomial RV's of order n and m with the same success probability p.

A Binomial RV X of order n represents the number of successes in n Bernoulli independent trials and as such can be seen as a sum of n binary independent RV's V_i whose distribution is

$$P(V_i = 1) = p \qquad P(V_i = 0) = q = 1 - p$$

where p is the probability of a successful trial. It follows that the sum of two V.C. Independent binomial of order h and k with the same success probability p is still a binomial of order $k + h$, and, therefore, also the Binomial family is closed respect to the operation of convolution.

It is left to the reader to verify the above by expressly applying the convolution operation.♣

By the definition we see that the average function is interchangeable with a linear operation and, therefore, we have

$$E[X+Y] = E[X] + E[Y] \tag{3.25}$$

We have also, and the proof is left to the reader,

$$\mathrm{VAR}[X+Y] = \mathrm{VAR}[X] + \mathrm{VAR}[Y] + \mathrm{COVAR}[XY] \tag{3.26}$$

Of course, RVs are independent, or just uncorrelated, we have

$$\mathrm{VAR}[X+Y] = \mathrm{VAR}[X] + \mathrm{VAR}[Y] \tag{3.27}$$

In particular, the variance of the average of n values, $\overline{X}_n$, is

$$\mathrm{VAR}[\overline{X}_n] = \frac{\mathrm{VAR}[X]}{n} \tag{3.28}$$

which explains the law of large numbers, as the variance of $\overline{X}_n$ decreases as $n \to \infty$

3.5 More general functions

In general we can have more functions of more RVs.

$$Z = g(X,Y) \qquad W = h(X,Y) \tag{3.29}$$

We can show that the joint densities of the two pairs of RV are related by the

Theorem: (3.30)
We have

$$f_{zw}(z,w) = \sum_{n=1}^{\infty} \frac{f_{XY}(x_n, y_n)}{\mid J(x_n, y_n) \mid} \tag{3.31}$$

where $(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n) \ldots$ represent the rela solution of the system of equations

$$\begin{cases} g(x,y) = z \\ h(x,y) = w, \end{cases} \tag{3.32}$$

and $J(x,y)$ is the Jacobian of transformtion (3.32),

$$J(x,y) = \begin{vmatrix} \dfrac{\partial g(x,y)}{\partial x} & \dfrac{\partial g(x,y)}{\partial y} \\ \\ \dfrac{\partial h(x,y)}{\partial x} & \dfrac{\partial h(x,y)}{\partial y} \end{vmatrix} \tag{3.33}$$

Example (3.34)

Let X and Y continuou RVs with Joint cdf $F_{XY}(x,y)$. If we adopt the transformations

$$\begin{cases} z = F_X(x) \\ w = F_Y(y/x) \end{cases} \tag{3.35}$$

which presents the unique solution

$$\begin{cases} x_1 = G(z) & (0 \leq z < 1) \\ y_1 = H(w,z) & (0 \leq w < 1) \end{cases} \tag{3.36}$$

The Jacobian is

$$J(x,y) = f_X(x) f_Y(y/x) = f_{XY}(x,y)$$

e dalla (3.31)

$$f_{ZW}(z,w) = \frac{f_{XY}(x_1,y_1)}{\mid J(x_1,y_1) \mid} = 1. \qquad (0 \leq z, w < 1)$$

Therefore Z e W are uniform and independent in $[0;1]$.

The above provides a generalization of we have seen in Theorem 3.4, and shows how to get two RVs of any joint distribution starting from two uniform and independent RVs.

3.6 Problems for solution

Chapter 3

P.3.1 f_X, f_Y are uniform within intervals respectively $[0;5]$ $[-3,-1]$. Find the pdf of RVs' (6.1)

1. $Z = X + Y$
2. $W = X - Y$

P.3.2 Let X e Y be independent RVs' with negative exponential pdfs' and average value $\frac{1}{\lambda}$. Find the pdf of RVs (6.2)

1. $Z = X - Y$
2. $W = X + \dfrac{Y}{2}$

P.3.3 Find $P(Z = n)$ where $Z = X + Y$ is the sum of the numbers that appear in the rolling of two dices. (6.6)

Solutions of Problems on Probability theory

Chapter 1

P.1.1 Rolling three dices, evaluate the probability of having k equal faces, with $k \in [0; 2; 3]$.

Solution.

$$P(0) = \frac{(6)_3}{6^3} = 0,\overline{5} \qquad P(2) = \frac{6 \cdot 3 \cdot 5}{6^3} = 0,41\overline{6} \qquad P(3) = \frac{6}{6^3} = 0,02\overline{7}$$

P.1.2 Rolling a dice three times, evaluate the probability of having at least one 6.

Solution.

$$p = 1 - \frac{5^3}{6^3} = 0,42129..$$

P.1.3 Assuming women and men exist in equal number, and assuming that 5% of the men are colour blind and that 0,25% of the women are colour blind, evaluate the probability that a person drawn at random is colour blind. Then evaluate the probability that, having drawn a colour-blind person, this is a male.

Solution. By the total probability theorem

$$P(D) = \frac{1}{2}\frac{5}{100} + \frac{1}{2}\frac{0.25}{100} = 0.02625$$

Using Bayes Theorem

$$P(M|D) = P(D|M)\frac{P(M)}{P(D)} = \frac{5}{100}\ \frac{1/2}{0.02625} = 0.952$$

P.1.4 Drawn a card from a deck of 52 cards, verify wheter the following events are statistically independent:

a) $A = \{\textit{drawing of a picture card}\};\qquad B = \{\textit{drawing of a hearth card}\}$

b) What if the king of hearths is missing from the deck of cards?
c) What if a card, at random, is missing?

Solution. a) independent b) not independent c) independent

P.1.5 A dice A has four red faces and two white faces. A dice B, vice-versa, has two red faces and four white faces. You flip a coin once, if heads the game continues with dice$A$, otherwise it continues with dice B. a) On rolling the dice, what is the probability that a red face appears on the dice? b) and at the second rolling of the same dice? c) If the first two rollings show a red face, what is the probability that also on the third rolling is red? d) If the first n rollings show a red face, what is the probability that you are using dice A?

Solution.

$$a)\ \frac{1}{2} \qquad b)\ \frac{1}{2} \qquad c)\ \frac{3}{5} \qquad d)\ \frac{2^n}{1+2^n}$$

P.1.6 An urn contains two white balls and two black. A ball is drawn and replaced with a ball of a different colour. Then a second ball is drawn. Calculate the probability p that the first extracted was white, when the second is white.

Solution: $1/4$

P.1.7 The probabilities that three different archers, A, B hit the mark, independently of one another, are respectively $1/6$, $1/4$ and $1/3$. Everyone shoots an arrow. a) Find the probability that only one hits the mark. b) If only one hits the mark, what is the probability he is archer A?

Solution:

$$a) \quad \frac{31}{72} \qquad b) \quad \frac{6}{31}$$

P.1.8 A duel among three people A, B and C is carried out according to the Russian roulette. A six round revolver is loaded with two cartridges. The duelists pass cyclically the weapon, spinning the cylinder every time (so that each duelist has 1/3 probability of being on a loaded chamber) and shooting themselves as long as only one remains alive. Assuming that A is the first, what is the probability that each duelist is the first to die? b) and to win?

Solution:

1. $$p_A = \frac{9}{19} \qquad p_B = \frac{6}{19} \qquad p_C = \frac{4}{19}$$
2. $$p_A = \frac{56}{209} \qquad p_B = \frac{69}{209} \qquad p_C = \frac{84}{209}$$

You have at least two ways to get at the result. The one is to evaluate all the sequence. For example, if p_A is the probability that A dies first we have

$$p_A = 1/3 + (2/3)^3(1/3) + (2/3)^6(1/3) + ... = (1/3)\sum_{i=0}^{\infty}(8/27)^i = 9/19$$

The second one is to observe that

$$p_B = (2/3)p_A; \qquad p_C = (2/3)^2 p_A.$$

Since we have also

$$p_A + p_B + p_C = 1$$

this provide an equation whose solution is p_A, and the others follow.

From a deck of 52 cards we draw two cards. Find the probabilities of the following events $A = \{$the first card is a King; the second figure$\} = \{K_1; F_2\}$, $\quad B = \{$at least one figure$\}$, $C = \{$the second card is a King$\}$.

Solution: We can count favorable outcomes in the space of outcomes, represented by the pair of card. Alternatively, we can use the conditional probability definition:

$$p(A) = P(K_1; F_2) = P(K_1)P(F_2|K_1) = \frac{4}{52}\,\frac{11}{51} = 0,0165\dots$$

$$p(B) = \frac{12}{52} + \frac{12}{52} - \frac{12}{52}\,\frac{11}{51} = 0,411\dots$$

$$p(C) = \frac{4 \times 51}{52 \times 51} = \frac{4}{52}.$$

Note that, to some, result $p(C)$ is unexpected. This is because of a wrong interpretation of the role of time. Time has no effect whatsoever on the event. The fact that we refer to an event before or after another is only to add flavor to the example.

Chapter 2

P.2.1 Given the function $f(x) = \frac{C}{\alpha^2 + x^2}$, determine the relationship between C e α in order to make $f(x)$ a pdf. (Cauchy). (3.1)

Solution. $\pi C = \alpha$

P.2.2 A point P uniformly chosen in a square of Side L centered at the origin and the x-axis. Find the pdf of RV X, coordinate of the orthogonal projection of P on the horizontal axis.

Solution.

$$f_X(x) = \frac{1}{L}, \qquad -L/2 \le x \le L/2$$

P.2.3 A point P uniformly chosen in a circle of radius R centered at the origin and the x-axis. Find the pdf of RV X, coordinate of the orthogonal projection of P on the horizontal axis.(3.2)

Solution.

$$f_X(x) = \frac{2}{\pi R}\sqrt{1 - (\frac{x}{R})^2}, \qquad -r/2 \le x \le R/2$$

P.2.4 Find the first order moment of pdf $f(x) = \lambda^2 x e^{-\lambda x}$, $x \ge 0$, and 0 elsewhere.

Solution. $2/\lambda$.

P.2.5 Find the first order moment of the integer distributions

1. $P(X = k) = (1-p)^{k-1}p, \quad k \ge 1;$
2. $P(X = k) = (1-p)^k p, \quad k \ge 0.$

Solution. $1/p$, and $(1-p)/p$.

P.2.6 2 points are chosen uniformly and independently in a segment of length L. Find the pdf of RV X distance to the origin of the point closest to the origin. Find the joint pdf of (X, Y) where Y is distance to the origin of the point farthest to the origin. Extend the result to the case of n points.(3.6)

Solution.
As usual

$$P(x < X \le x + \Delta x) = P(\text{one point in } \Delta x\text{; the other beyond x})$$

We get

$$f_X(x) = \frac{2}{L}\frac{L-x}{L}, \qquad 0 \le x \le L$$

For the joint pdf we have

$$P(x < X \le x + \Delta x;\ y < Y \le y + \Delta y) = 2\frac{\Delta x}{L}\frac{\Delta y}{L}, \qquad x < y$$

$$f_{XY}(x, y) = \frac{2}{L^2}, \qquad x < y$$

P.2.7 2 points are chosen uniformly and independently in a circle of radius R. Find the pdf of RV X distance to the center of the point closest to the center.(3.8)

Solution.

$$f_X(x) = 2n \, \frac{x}{R^2} \left[1 - (\frac{x}{R})^2\right]^{n-1} \qquad (0 \leq x \leq R)$$

P.2.8 Take a number X from one to six, throw three dices. You win C if X appears once, $2C$ if X appears twice, $3C$ if it appears three times, and you lose C if X does not appear. Check whether this is a fair game. (3.12)

Solution.
The number of possible outcomes is $6^3 = 216$. X appears once with probability $3 \times 25/216$, twice with probability $3 \times 5/216$, and thrice with probability $1/216$. The probability of winning is the sum, i.e, $91/216$; therefore your loss is on the average $(125/216)C$. On the other side, if your win is, on the average,

$$(75/216)C + (15/216)2C + (1/216)3C = (108/216)C$$

Then on the average you loose $(17/216)C$ at each bet.

P.2.9 Assume the RV X, lifespan of a component, is uniform in $[0; L]$. We know that the component age is z; find the pdf of its lifespan. Find the pdf of Y, remaining lifespan.

Solution.
We look for

$$f_X(x|\text{age=z}) = f_X(x|X > z) = \lim \frac{P(x < X \leq x + \Delta x; X > z)}{\Delta x P(X > z)} =$$

$$= \lim \frac{P(x < X \leq x + \Delta x)}{\Delta x P(X > z)}, \qquad x > z$$

Therefore

$$f_X(x|X > z) = \frac{f_X(x)}{P(X > z)} = \frac{1}{L - z}, \qquad z \leq x \leq L$$

Then

$$f_Y(y) = f_X(y + z|X > z) = \frac{1}{L - z}, \qquad 0 \leq x \leq L - z$$

P.2.10 Repeat the previous exercise assuming that the pdf of X is negative exponential. Find the fair amount a a customer of age z must pay to get a capital C if he dies before the year.

Solution.
As in the previous case we have

$$f_X(x|X > z) = \frac{f_X(x)}{P(X > z)} = \frac{\lambda e^{-\lambda x}}{e^{-\lambda z}}, \qquad x > z$$

$$f_Y(y) = f_X(y+z|X>z) = \frac{\lambda e^{-\lambda(y+z)}}{e^{-\lambda z}} = \lambda e^{-\lambda y}, \qquad x > 0$$

Note that the remaining life span has still the same pdf as the original lifespan. This is because the negative exponential is "memoryless".

The probability of dying within a year is $P(X < 1) = 1 - e^{-\lambda}$. The fair amount a is such that

$$ae^{-\lambda} = C(1 - e^{-\lambda})$$

P.2.11 Check whether functions of x and y below can represent joint pdfs' and if so check whether X and Y are statistically independent. (5.1)

1. $f(x,y) = 4xy \quad (0 \le x \le 1;\ 0 \le y \le 1),$
2. $f(x,y) = 8xy \quad (0 \le x \le y;\ 0 \le y \le 1),$
3. $f(x,y) = 4x^2y \quad (0 \le x \le 1;\ 0 \le y \le 1)$

P.2.12 A person in phone booth makes a phone call whose duration is represented by RV X, with negative exponential pdf with mean value $1/\mu$. A second person comes after a time Y. RV negative exponentially with average $1/\lambda$, independent of X. Find the pdf of RV W, the time the latter has to wait to the end of the call. (5.6)

Solution.
If $Y > X$ the second person arrives when the first has already finished his phone call and, therefore $W = 0$. On the other side, we take the condition $Y = y, Y < X$. RV W is then "the remaining lifespan" of Problem 2.10. By this problem we have learned that, with the negative exponential pdf, the remaining lifespan has the same pdf. Therefore

$$f_W(w|Y = y; Y < X) = \mu e^{-\mu w}, \qquad y \ge 0$$

that doesn't depend on y. The, using the total probability Theorem

$$f_W(w|Y < X) = \mu e^{-\mu w}, \qquad y \ge 0$$

Finally

$$f_W(w) = \delta(w)P(Y > X) + \mu e^{-\mu w}P(Y < X), \qquad y \ge 0$$

See Problem 2.13 to see $P(Y > X)$.

P.2.13 Given two independent RVs' X, and Y, find the probability of the event $\{Y \le X\}$ when

1. f_X, f_Y are uniform within intervals respectively $[-1; 3], [0; 4]$;
2. f_X, f_Y with the same pdf (you do not need to know the pdf);
3. f_X, f_Y are negative exponentials with parameters λ and μ;

What about event $\{Y \leq X/2\}$?

Solution.
The first way finds event $\{Y \leq X\}$ in the plane x, y, and then integrates the joint pdf in such event. Case 1) is very simple since the joint distribution is uniform and the integration takes the volume of the regular solid where the base is the area of the event. The area of the vent is the portion of the rectangle of base $[-1; 3]$ and height $[0; 4]$ that lies beneath straight line $y = x$. The height of the pdf is $1/16$.

The second way uses conditioning.

$$P(Y \leq X) = \int P(Y \leq X|X = x) f_X(x) dx$$

In the first case we have

$$P(Y \leq X|X = x) = x/4, \qquad 0 \leq x \leq 3$$

$$P(Y \leq X|X = x) = 0, \qquad -1 \leq x \leq 0$$

$$P(Y \leq X) = \int_0^3 x/4 \; 1/4 dx = 9/32$$

In case 2, since both RVs' obeys the same law, outcomes $X > Y$ or $Y > X$ are equally probable.

In case 3 we use conditioning again.

$$P(Y \leq X) = \int P(Y \leq X|Y = y) f_Y(y) dy = \int_0^\infty P(X \geq y|Y = y) f_Y(y) dy =$$

$$= \int_0^\infty e^{-\mu y} \lambda e^{-\lambda y} dy = \frac{\lambda}{\lambda + \mu} \int_0^\infty (\lambda + \mu) e^{-(\lambda+\mu)} dy = \frac{\lambda}{\lambda + \mu}$$

P.2.14 Find the pdf of RV $Z = \min(X, Y)$, where X and Y are two independent negative exponential RVs' with parameters λ and μ respectively. (Hint: observe that $\min(X, Y) > z$ if $x > z$ and $Y > z$. Also, we may take the condition $Y = y$...)

Solution.
The suggestion says

$$P(Z > z) = P(\min(X, Y) > z) = P(X > z; Y > z) = P(X > z) P(Y > Z) = e^{-\lambda z} e^{-\mu z}$$

Therefore

$$P(Z > z) = e^{-(\lambda+\mu)}$$

or

$$f_Z(z) = (\lambda + \mu) e^{-(\lambda+\mu) z}, \qquad z \geq 0$$

Take notice: The minimum of two negative exponential RVs is again a negative exponential RV with parameters sum of parameters.

P.2.15 Take interval $[0, X]$, where X is a RV Erlang-2. Then take a point P uniformly within the preceding interval. Find the pdf of Y, length of $\overline{0P}$.

Solution.
We use the Total Probability Theorem

$$f_Y(y|X = x) = \frac{1}{x}, \qquad y \le x$$

$$f_Y(y) = \int f_Y(y|X = x) f_X(x)dx = \int_y^\infty \frac{1}{x}\lambda^2 x e^{-\lambda x}dx = \lambda \int_y^\infty \lambda e^{-\lambda x}dx =$$

$$= \lambda(1 - F_X(y) = \lambda e^{-\lambda x}$$

The solution could be expected. Why?

P.2.16 n points are uniformly taken within $[0; T]$. Find the probability that k out of n point lie within an interval $[0; X]$ where RV X is uniform in $[0; T]$.

Solution.
The probability can be written as:

$$P(N(X) = k/N(T) = n) = \int_0^T P(N(X) = k/N(T) = n, X = x) f_X(x)dx =$$

$$= \int_0^T \left(\begin{array}{c} n \\ k \end{array} \right) \left(\frac{T-x}{T}\right)^{n-k} \left(\frac{x}{T}\right)^k \frac{1}{T}dx = \frac{1}{n+1}$$

It seems strange that the solution does not depend on k. This becomes apparent if we solve the problem in this other way. The extreme of interval X is itself a uniform point in $[0, T]$, exactly as the others n. Therefore, the sought probability is the probability that this boundary point lies the $k+1$-th position out of $n+1$. But all positions are equally probable and therefore the sought probability is $1/(n + 1)$.

P.2.17 Two RVs' X and Y are independent and uniformly distributed in $[0; 1]$. Find $f_X(x|X > Y)$, $f_{XY}(x, y|X > Y)$ and $P(X > 2Y|X > Y)$.

Solution.

$$f_X(x|X > Y) = P(X > Y|X = x)\frac{f_X(x)}{P(X > Y)} = x\frac{1}{1/2} = 2x, \qquad 0 \le x \le 1$$

$$f_{XY}(x, y|X > Y) = P(X > Y|X = x, Y = y)\frac{f_{XY}(x, y)}{P(X > Y)} = 2 \qquad x > y$$

Since the above is uniform in $0 \le y \le x \le 1$, $P(X > 2Y|X > Y)$ is simply the ratio of the areas of events $X > 2Y$ and $X > Y$, equal to $1/2$.

Chapter 3

P.3.1 f_X, f_Y are uniform within intervals respectively $[0;5]$ $[-3,-1]$. Find the pdf of RVs' (6.1)

1. $Z = X + Y$
2. $W = X - Y$

Solution.

$$f_z(z) = \begin{cases} (1/10)(z+3) & -3 \leq z \leq -1 \\ 1/5 & -1 \leq z \leq 2 \\ (1/10)(4-z) & 2 \leq z \leq 4 \end{cases}$$

$$f_W(w) = \begin{cases} (1/10)(z-1) & 1 \leq w \leq 3 \\ 1/5 & 3 \leq w \leq 6 \\ (1/10)(8-w) & 6 \leq w \leq 8 \end{cases}$$

P.3.2 Let X e Y be independent RVs' with negative exponential pdfs' and average value $\frac{1}{\lambda}$. Find the pdf of RVs (6.2)

1. $Z = X - Y$
2. $W = X + \dfrac{Y}{2}$

Solution.

$$a) \quad f_Z(z) = \frac{1}{2}\lambda e^{-\lambda|z|} \quad \text{(Laplace)}$$

$$b) \quad f_W(w) = 2\lambda(e^{-\lambda z} - e^{-2\lambda z}) \qquad (z > 0)$$

P.3.3 Find $P(Z = n)$ where $Z = X + Y$ is the sum of the numbers that appear in the rolling of two dices. (6.6) **Solution.**

$$P(Z = n) = \begin{cases} \dfrac{n-1}{36} & (2 \leq n \leq 7) \\ \\ \dfrac{13-n}{36} & (7 \leq n \leq 12) \end{cases}$$

F. Borgonovo

Class notes on Traffic Theory

Contents

Chapter 1

Discrete-state processes and Markov Chains

1.1 Introduction

Many of the disciplines that are part of the Telecommunications area have resulted in the development of the theory of *Stochastic Processes* or *Random Processes*. In this theory, the concepts of the field of random variables are extended to include the time variable, so as to be able to represent random phenomena that vary over time. In particular, in the field of Telecommunication Networks, the advent of the telephone at the end of the nineteenth century has immediately forwarded the problem of the sizing of telephone switches and the respective connections (trunks). Such sizing must make use of probabilistic tools, as the service request is a variable that must be addressed with probabilistic tools. The theory that emerged has been developed with applications to modern telecommunications networks, but also with applications in many other areas of science, including biology and even economics.

Given the close connection between theory and applications, below these two issues will be developed together, introducing theoretical concepts and illustrating their impact on particular applications involving telecommunications networks. This has been performed in the hope to make the theory more easily understood and its use in applications clearer.

The theoretical concepts set out below make use of concepts of the *Elementary Theory of Random Variables*, which are therefore considered to be known to the reader.

1.2 Definition of Random Processes

The concept of random process is an extension of the concept of Random Variable. Specifically, given an experiment ξ, a result set $S = \{\alpha_i\}$ (the universe), and a probability measure $P_S(.)$ on subsets of S, we define the random process associated with ξ as a real function $X(t, \alpha)$ defined on the space S so that each result $\alpha_i \in S$ of ξ corresponds to a real function of time $X(t, \alpha_i) = x_i(t)$

defined on a set of the real axis t (figure 1.1). In this way each subset A of results α is made to correspond to a subset of functions $x_i(t)$ t on the real axis, each function being called a *sample* of the process (Figure 1.2).

We note that if time t is considered as a constant, the definition above reduces to a Random Variable (RV), a concept we assume as known.

The classes of functions associated with a process can be of many types, analytically defined or not, continuous or discontinuous, differentiable, etc.

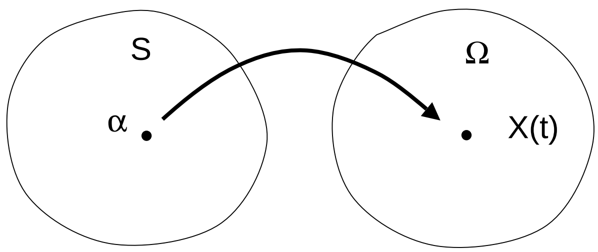


Figure 1.1: *Process mapping.*

A first important characterization distinguishes if t is continuous, and we say continuous-time processes, or discrete ($t = 1, 2, \ldots, n \ldots$), and we say say discrete-time processes.

A second characterization relates to the definition set of of $x(t)$ which can be continuous or discrete ($x = x_1, x_2, \ldots, x_n, \ldots$). The latter case gives rise to discrete state processes or *Chains.*

A given Random Process $X(t, \alpha)$ sampled at a precise instant $t = t_0$, provides a Random Variable (RV) $X(t_0) = X(t_0, \alpha)$, whereas in time instants $t_1, t_2, \ldots, t_n$ it provides a succession of RV's $X(t_1), X(t_2), \ldots, X(t_n)$, described below with the reduced notation $X_1, X_2, \ldots, X_n$. These sequences for *all* $t_1, t_2, \ldots, t_n$, *and for all* n, are referred to as the *family of the RV's of the process.*

It can be shown that the description of a process, i.e. the description that allows to evaluate any probability of events related to the process, is given by the description the family of RV's of the process, the description being given by various forms introduced for the description of random variables, for example the *Cumulative distribution function* (c.d.f.):

$$\begin{aligned} &F_{X_1X_2\cdots X_n}(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n) = \\ &\qquad = P(X(t_1) \le x_1, X(t_2) \le x_2, \ldots, X(t_n) \le x_n) \end{aligned} \tag{1.1}$$

or the *probability density function* (pdf), useful for continuous space RV's,:

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \frac{\partial^n F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)}{\partial x_1\, \partial x_2\, \ldots\, \partial x_n} \tag{1.2}$$

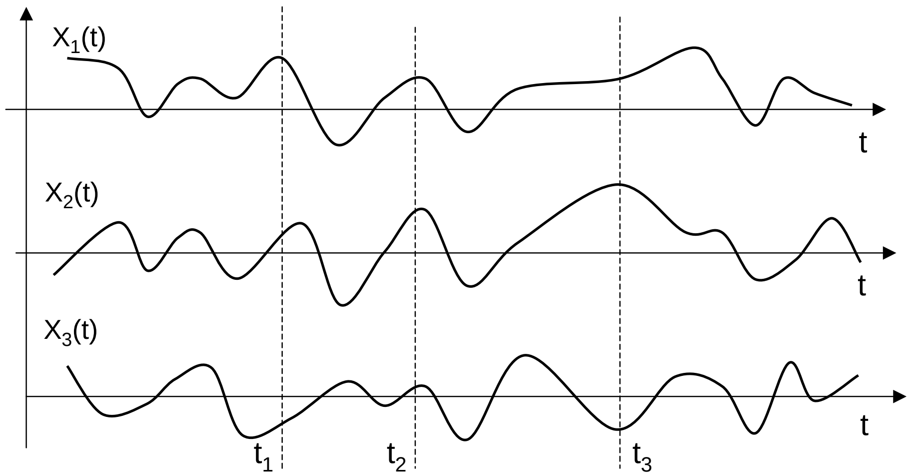


Figure 1.2: *Process samples.*

or the *probability distribution function* (p.d.f.) for discrete space RV's:

$$P(X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n) \tag{1.3}$$

To avoid too heavy a notation for distributions, when there is no danger of misinterpretation, we will simply denote by X_k the event $X_k = j_k$, and therefore the distribution of order n seen above can be expressed as

$$P(X_0, X_1, X_2, \ldots X_n) \tag{1.4}$$

1.3 Purely Random Processes

Purely Random processes are those for which all the variables $X_0, X_1, \ldots X_n$, for all n are statistically independent; which means that the joint probability of each string of n variables, with any n, splits into the product of the corresponding probabilities:

$$P(X_{j_1}, X_{j_2}, \ldots X_{j_{n-1}}, X_{j_n}) = P(X_{j_1})P(X_{j_2}), \ldots P(X_{j_{n-1}})P(X_{j_n}),$$

for any sequence $X_{j_1}, X_{j_2}, \ldots X_{j_{n-1}}, X_{j_n}$, for any n.

The simple nature of these processes makes them unsuitable for modeling many processes. However, there is the case of *Arrival Processes* in which Purely Random Processes provide useful models.

1.3.1 Bernoulli Arrivals

This process describes single *arrivals*, on the discrete-time axis and is defined by three *axioms*. Denoted by $A(n)$ the binary RV Arrivals at time n, we have:

$$P(A(n) = 1) = p, \qquad \forall n. \tag{1.5}$$

This is called the *stationarity* axiom

$$P(A(n) > 1) = 0 \qquad \forall n. \tag{1.6}$$

This is called the *singularity* axiom

The third axiom is the *independence* axiom, and states that the RV's of any family $A(i), A(j), \ldots, A(k)$, are statistically independent. For example we have

$$P(A(n_1) = x_1, A(n_2) = x_2) = P(A(n_1) = x_1)P(A(n_2) = x_2), \qquad x_1, x_2 \in [0, 1]. \tag{1.7}$$

From the definition above, denoted by $A(n, n+m-1)$ total the number of arrivals in the interval $[n, n+1, \ldots, n+m-1]$, composed of m instants,

Theorem: (1.8)
the distribution of $A(n, n+m-1)$ *is given by*

$$P(A(n, n+m-1) = k) = \left(\begin{array}{c} m \\ k \end{array} \right) p^k (1-p)^{m-k} \qquad 0 \le k \le m \tag{1.9}$$

This comes from the number of ways in which the k arrivals can be chosen among the m instants, being each possible pattern equally probable.

The above distribution is known as the *Binomial Distribution*, with

$$E[A] = mp \tag{1.10}$$

$$\mathrm{VAR}[A] = mp(1-p), \tag{1.11}$$

it is bell shaped with a maximum close to, or coincident with, mp.

We also have

Theorem: (1.12)
The distance X_i between consecutive arrivals i and $i+1$ has distribution

$$P(X = k) = p(1-p)^{k-1} \qquad k \ge 1 \tag{1.13}$$

This holds because we must have $k-1$ failures and success (arrival) in choosing the arrival point.

The above distribution is known as the *Geometric Distribution*, with

$$E[X] = 1/p \tag{1.14}$$

$$\text{VAR}[X] = (1/p)^2, \tag{1.15}$$

and is an ever decreasing function.

Example *Random Access-1-Finite population* (1.16)
Data packets of a user arrive according to Bernoulli arrivals with probability p at each time (slot). Consider n such users and assume that at each arrival the user transmit its packets. If more than one user is transmitting at the same instants packets are destroyed (collided) and are cleared, the other are successful. Find the average number per slot that are successful.

The probability that k users are transmitting at the same instant is the Binomial distribution with parameters n and p. The he average number of successful packets/slot (there can be only one or zero), is called throughput S and coincides with the probability that there is only one packet being transmitted, i.e.,

$$S = np(1-p)^{n-1}.$$

We express the above as function of the channel traffic G, defined as the average number of transmitted packets at each time, i.e., $G = np$. We have

$$S = G(1 - \frac{G}{n})^{n-1}. \tag{1.17}$$

Note that we also have

$$\lim_{n\to\infty} S = Ge^{-G}. \tag{1.18}$$

In any cases s is maximized when $G = 1$, where in the limit we get a throughput equal to $e^{-1} = 0.367...$ This optimal setting setting is possible only knowing n so that we can adopt $p = 1/n$.

1.3.2 Arrivals in continuous time

Here we consider arrivals on the time axis that have a finite density, which means that two consecutive arrivals are separated by a non-zero interval on the average.

Let $A(t) = A(t, t + \Delta t)$ be the number of arrivals in the interval $[t, t + \Delta t]$, where we assume that Δt is very small, i.e., $\Delta t \to 0$. If we exclude multiple arrivals within the same Δt, this is a binary process that is 1 within the interval Δt in which an arrival occurs and is zero elsewhere (Fig. 1.3).

The average arrival frequency in $[t, t + \tau]$ is

$$\nu(t, t+\tau) \triangleq \frac{E[A(t, t+\tau)]}{\tau}, \tag{1.19}$$

whereas the instantaneous frequency, or density, or rate, at time t is

$$\lambda(t) \triangleq \lim_{\Delta t\to 0} \nu(t, t+\Delta t) = \lim_{\Delta t\to 0} \frac{E[A(t, t+\Delta t)]}{\Delta t} \tag{1.20}$$

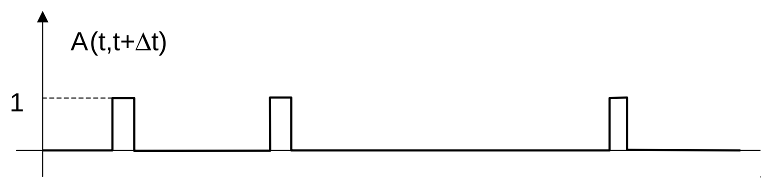


Figure 1.3:

For small Δt we have

$$E[A(t,t+\Delta t)] = \lambda(t)\Delta t + \omega(\Delta t) \tag{1.21}$$

where $\omega(\Delta t)$ denotes terms of infinitesimal order higher than Δt.

If we want to exclude multiple arrivals in the same time instant we must have

$$P(A(t,t+\Delta t) > 1) = \omega(\Delta t) \tag{1.22}$$

which shows that $A(t,t+\Delta t)$ is in the limit a binary variable and that $E[A(t,t+\Delta t)] = P(A(t,t+\Delta t) = 1)$. This, from (1.21) provides us with the following interpretation on the density $\lambda(t)$:

$$P(A(t,t+\Delta t) = 1) = \lambda(t)\Delta t + \omega(\Delta t) \tag{1.23}$$

or

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(A(t,t+\Delta t) = 1)}{\Delta t} \tag{1.24}$$

In the following, in order to simplify notation, the term $\omega(\Delta t)$ is omitted.

1.3.3 Poisson Arrivals

Poisson Arrivals represent the counterpart of Bernoulli Arrivals in continuous time. Axioms (1.5), (1.6) and (1.7) become

$$P(A(t,t+\Delta t) = 1) = \lambda\Delta t + \omega(\Delta t), \qquad \text{stationarity} \tag{1.25}$$

$$P(A(t,t+\Delta t) > 1) = \omega(\Delta t) \qquad \text{singularity} \tag{1.26}$$

and, again, the *independence* axiom, that states that the RV's of any family $A(i), A(j), \ldots, A(k)$, are statistically independent. For example, for disjoint intervals, we have

$$\begin{aligned} & P(A(t_1, t_1 + \Delta t) = x_1, A(t_2, t_2 + \Delta t) = x_2) = \\ = & P(A(t_1, t_1 + \Delta t) = x_1) P(A(t_2, t_2 + \Delta t) = x_2), \qquad \forall x_1, x_2 \in [0, 1]. \end{aligned} \tag{1.27}$$

For the arrivals in some interval $[t; t + \Delta t]$ we have

Theorem: (1.28)
The distribution of the arrivals in $[t, t + \tau]$ *is*

$$P(A(t, t + \tau) = k) = \frac{(\lambda \tau)^k}{k!} \, e^{-\lambda \tau} \qquad k \geq 0 \tag{1.29}$$

Proof
Let subdivide the interval τ in $n = \tau / \Delta t$ small intervals Δt. If $\Delta t \to 0$ (and $n \to \infty$), we can have at most one arrival in Δt and arrivals become Bernoulli with $p = \lambda \Delta t$. Then we have

$$\begin{aligned} P(A(\tau) = k) &= \lim_{\Delta t \to 0 \; n \to \infty} \binom{n}{k} (\lambda \Delta t)^k (1 - \lambda \Delta t)^{n-k} = \\ &= \lim_{n \to \infty} \frac{n!}{k!(n-k)!} \left(\frac{\lambda \tau}{n} \right)^k \left(1 - \frac{\lambda \tau}{n} \right)^{n-k} = \\ &= \frac{(\lambda \tau)^k}{k!} \lim_{n \to \infty} \left(\frac{n}{n} \, \frac{n-1}{n} \, \ldots \, \frac{n-k+1}{n} \right) \left(1 - \frac{\lambda \tau}{n} \right)^{n-k} \end{aligned}$$

By the fundamental limit

$$\lim_{x \to \infty} \left(1 + \frac{1}{x} \right)^x = e$$

for finite values of k we get (1.29) ♣.

Distribution (1.29) is called the *Poisson Distribution*, with

$$E[A(\tau)] = \lambda \tau \tag{1.30}$$

$$\mathrm{VAR}[A(\tau)] = \lambda \tau. \tag{1.31}$$

It is bell shaped with a maximum close to, or coincident with the average.

Note that the set of measure τ need not necessarily be an interval; in fact it can be obtained also as union of disjoint intervals having overall measure τ. For example, if we consider two disjoint intervals of measure τ_1 and τ_2 with $\tau = \tau_1 + \tau_2$, we have

$$A(\tau) = A(\tau_1) + A(\tau_2).$$

According to the probability theory, the distribution of A is the convolution of the distributions of A_1 and A_2. On the other side, being all the involved distributions of the Poisson type we can conclude that the convolution of two or more Poisson distribution is still Poisson distributed with average equal to the sum of the average of the components.

Theorem: (1.32)
The distance X_i between consecutive arrivals i and $i+1$ has pdf

$$f_X(x) = \lambda e^{-\lambda x} \qquad x \geq 0 \tag{1.33}$$

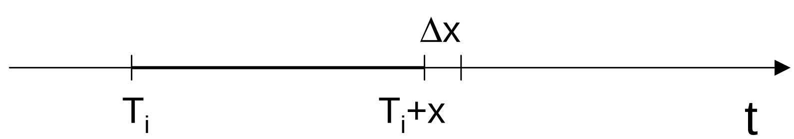


Figure 1.4:

This comes straight down from the definition of pdf $f_X(x)$:

$$f(x)\Delta x = P(x < X \leq x + \Delta x) = P(N(x) = 0)P(N(\Delta x) = 1) = \; e^{-\lambda x}\lambda \Delta x$$

X_i e X_j, $i \neq j$ are statistically independent because referring to disjoint intervals.♣

Pdf (1.33) is called *negative exponential* and is determined by the lone parameter λ and is such that

$$E[X] = 1/\lambda, \qquad \text{Var}[X] = E[X]^2. \tag{1.34}$$

It is left to the reader to show that the distance X_{ik} between arrivals i and $i+k$ has pdf

$$f_X(x) = \frac{(\lambda x)^{k-1}}{(k-1)!}\,\lambda\, e^{-\lambda x} \qquad x \geq 0 \tag{1.35}$$

whose average is k/λ. Such a pdf is called Erlang-k.

Poisson arrivals have other simple properties:

Property *(Union)* (1.36)
The union of two independent Poisson Arrivals $A_1(t)$, $A_2(t)$ still represents Poisson Arrivals $A(t)$ with frequency equal to the sum of the frequencies λ_1 and λ_2 of the two component processes.

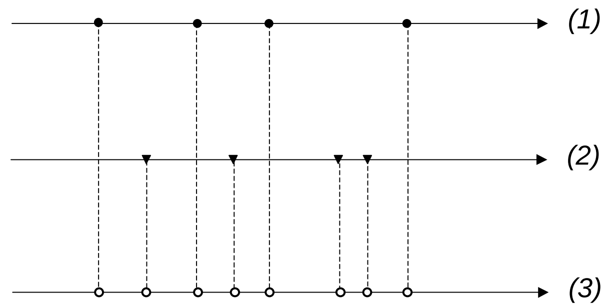


Figure 1.5: Union of two flows.

Property *(Random Decomposition)* (1.37)
The Arrival process $A_p(t)$ that is obtained from the Poisson Arrivals $A(t)$ by suppressing (or not) each arrival with probability $(1-p)$, all independently of each other, again represents Poisson Arrivals with average frequency equal to p times the original frequency Λ.

The proof of these properties is very simple since it only requires to verify that the Poisson axioms hold. For example, For the first axiom we have, in the two cases

$$P(A(t, t+\Delta t) = 1) = (\lambda_1 + \lambda_2)\Delta t + \omega(\Delta t),$$
$$P(A_p(t, t+\Delta t) = 1) = \lambda p \Delta t + \omega(\Delta t).$$

It is left to the reader to verify that also the two other Poisson axioms hold.

Example *(Deterministic decomposition)* (1.38)
The Arrival process $A_p(t)$ that is obtained from the Poisson Arrivals $A(t)$ by suppressing arrivals alternatively, e.g., those that have an odd order of arrival.

Here, an arrival in $[t, t+\Delta t]$ requires an arrival in the original process, which is $\lambda\Delta t$, but, whether the arrival is suppressed or not depends on what has happened in the past, and the third Poisson axiom does not hold.

Example *Random Access-2: Infinite population* (1.39)
Data packets of different users arrive according to Poisson arrivals with rate λ. Upon arrival each packet is transmitted, being T the transmission time of each packet. If more than one packet is being transmitted at the same time packets are destroyed (collided) and are cleared, the other are successful. Find the rate and the fraction of packets that are successful.

Consider a packet being transmitted within interval $[t, t+T]$. In order to avoid a collision, no packets must be transmitted within $[t-T, t+T]$, which occurs with probability

$$p_s = e^{-2\lambda T}. \quad (1.40)$$

Therefore, the rate of successful packets is

$$\lambda_s = \lambda e^{-2\lambda T}. \quad (1.41)$$

Note that (1.41) is maximized when $\lambda = 1/(2T)$, and, in this case we have

$$\lambda_s = 0.5e^{-1}. \tag{1.42}$$

Again, with this setting, also $p_s \to 0.5e^{-1}$, which is a too small fraction to be used in practice.

Example *Random Access-3: Infinite population* (1.43)
Referring to the previous example, assume that transmissions are synchronized at instants kT.

Consider a packet being transmitted within interval $[kT, (k+1)T]$. In order to avoid a collision, no packets must arrive within $[kT, (k+1)T]$, which occurs with probability

$$p_s = e^{-\lambda T}. \tag{1.44}$$

Therefore, the rate of successful packets is

$$\lambda_s = \lambda e^{-\lambda T}. \tag{1.45}$$

Again (1.41) is maximized when $\lambda = 1/(T)$, and, in this case we have

$$\lambda_s = \lambda e^{-1}. \tag{1.46}$$

1.4 Processes with memory

Any joint distribution of a finite sequence can be written as

$$P(X_0, X_1, \ldots, X_n, X_{n+1}) = P(X_0, X_1, \ldots, X_n) P(X_{n+1} | X_0, X_1, \ldots, X_n), \tag{1.47}$$

and iterating the conditioning we get

$$\begin{aligned} P(X_0, X_1, \ldots, X_n, X_{n+1}) = {} & P(X_0) P(X_1|X_0) P(X_2|X_0, X_1) \ldots \\ & \ldots P(X_n|X_0, X_1, \ldots, X_{n-1}), P(X_{n+1}|X_0, X_1, \ldots, X_n). \end{aligned} \tag{1.48}$$

This property shows how the joint description can be split into the product of descriptions of the single variables of the process, albeit conditional to the past history. This shows also how that samples of the process can be easily produced step by step, starting from the beginning at time $n = 0$. Once the sample is known in all past instants up to time n, the conditional distribution

$$P(X_{n+1}|X_0, X_1, \ldots, X_n) \tag{1.49}$$

shows how to get the sample value at n.

The dependance of (1.49) on states $X_0, X_1, \ldots, X_n$ is called *memory* and can be present at various degrees, while it is completely absent in Purely Random Processes, as seen in the preceding section.

Note that the decomposition expressed by (1.48) can be carried out in a variety of ways, for example starting with some value at $n+1$ and deriving past values by conditioning on the future. In this case we have memory of the future. Here, the reader must not be confused by the terms *past* and *future*, since this is only a conventional naming; we rather must think of the entire sequence $X_0, X_1, \ldots, X_n, X_{n+1}$ where we may know the value assumed by some variables, as shown in the next example. Conditioning as in (1.49), however, can reflect the way some processes are built with time in nature.

We also notice that, in general, the fact that past history is known does not make X_n independent of the future.

From each process $X(n)$ we can derive its *delayed* version $Y(n) = X(n-n_0)$ and its *reversed* version $Z(n) = X(n_0 - n)$. Once we know the description of $X(n)$, i.e., all the distributions of the family, we also know the description of $Y(t)$ and $Z(t)$ as follows

$$P(Y_1 = i_1, Y_2 = i_2, \ldots, Y_n = i_n) = P(X_{1-n_0} = i_1, X_{2-n_0} = i_2, \ldots, X_{n-n_0} = i_n),$$
$$P(Z_1 = i_1, Z_2 = i_2, \ldots, Z_n = i_n) = P(X_{n_0-1} = i_1, X_{n_0-2} = i_2, \ldots, X_{n_0-n} = i_n), \tag{1.50}$$

provided that $X(n)$ is defined at the required times.

1.5 Markov Chains

Markov Chains (MCs) (after the Russian mathematician A.A. Markov (1856-1922) represent the simplest kind of process other than the purely random processes. Referring to the discrete time axis they are defined by the property

$$P(X_{n+1}|X_0, X_1, \ldots X_n) = P(X_{n+1}|X_n), \qquad \forall X_0, X_1, \ldots X_n, X_{n+1}, \quad \forall n > 0 \tag{1.51}$$

This property shows that once we know the present state $X(n) = j_n$, then the past do not influence the occurrence of the future state $X(n+1)$. In other words, the experiment which determines X_{n+1} depends only on x_n, if known, and not from the past history $x_0, x_1, \ldots x_n$. *Note that this does not mean that X_{n+1} is independent from the past.* It becomes so only if X_n is known. By property (1.51) distribution (1.49) can be expressed as

$$P(X_0, X_1, \ldots, X_n, X_{n+1}) = P(X_0)P(X_1|X_0)P(X_2|X_1)\ldots P(X_n|X_{n-1}), P(X_{n+1}|X_n). \tag{1.52}$$

1.5.1 Properties of Markov Chains

Property (1.53)
Property (1.51) is symmetric in time. That is, once we know the present state, $X(n) = j$, then the knowledge of the future does not affect what is occurs in the past $X(n-1)$. (We remark, here, that the use of past, present and future has not a causal meaning. Present only means a chosen time n, past and future only mean, respectively, to the left and to the right of n. In this sense we may know the state of the process somewhere and not in the remaining part of the axis).

For simplicity we limit ourselves to show that

$$P(X_{n-1}|X_n, X_{n+1}) = P(X_{n-1}|X_n)$$

We apply Bayes' theorem to the probabilistic space conditional by the event $X(n) = j_n$, and we have

$$P(X_{n-1}|X_n, X_{n+1}) = P(X_{n+1}|X_n, X_{n-1})\frac{P(X_{n-1}|X_n)}{P(X_{n+1})|X_n)}.$$

The thesis follows immediately from the fact that, owing to (1.51), the first term in second member is equal to the denominator of the second term. ♣

The proofs of the following properties are easy and are left to the reader.

Property (1.54)
Once we know the present state, $X(n) = j$, then past and future are statistically independent, i.e.,

$$P(X_{n-1}, X_{n+1}|X_n) = P(X_{n-1}|X_n)P(X_{n+1}|X_n)$$

Property (1.55)
If we only know the state $X(m)$ of a chain at time m in the past, then future $X(n+1)$ does not depend on the past prior to m:

$$P(X_{n+1}|X_0, \ldots X_m) = P(X_{n+1}|X_m) \tag{1.56}$$

where $m < n$.

Relation (1.56) is often used as a definition of markov processes instead of (1.51), since it is more general.

The Markov property implies that all distributions of any order can be derived as products of distributions of the second and the first order.

Property (1.57)
The joint distributions of the third and higher orders can be derived from those of the second and first order.

This can be proved by (1.48) which, by the Markov property reduces to

$$P(X_0, X_1, \ldots, X_n, X_{n+1}) = P(X_0)P(X_1|X_0)P(X_2|X_1)\ldots\ldots P(X_n|X_{n-1}), P(X_{n+1}|X_n). \clubsuit \tag{1.58}$$

The above shows that a Markov Chain is entirely described by the *initial conditions*, $P(X_0)$ and the *second order conditional description*, $P(X_{n+1}|X_n)$, at any time $n \geq 0$. Then this description $P(X_{n+1}|X_n)$, called transition probability, is what, in fact, describe the way the process changes in time.

Property (1.59)
The transition probability of the reversed process is

$$P(X_{n-1}|X_n) = P(X_n|X_{n-1})\frac{P(X_{n-1})}{P(X_n)} \tag{1.60}$$

This is immediately derived by the definition of conditional probability.♣

The above shows that the transition probability of the reversed process does not depend on the lone transition probability of the original process, but also on the first order distribution, except when $P(X_n)$ does not depend on n.

1.5.2 Matrix representation of Markov Chains

Since the state space can be represented by integers, in many cases it is useful to introduce the matrix notation for distributions. With this notation

$$\mathbf{\Pi}(n) = \{\pi_i(n)\} = \{P(X(n) = i)\}$$

is the row vector that represents the first order probability distribution at time n, $(\pi_i(n))$. Since by (1.52) any joint distribution is expressed in terms of second order conditional distributions we introduce the matrices of these elements as

$$\mathbf{P}(n, n+m) = \{p_{jk}(n, n+m)\} = \{P(X(n+m) = k|X(n) = j)\}$$

The above provides *transition probabilities* from state j at time n to the generic state k at time $n+m$, whichever j, k are. The matrix itself is called the *m-step transition probability matrix* between n and $n+m$ and has the property that all its row sum to one; in fact, elements $p_{jk}(n, n+1)$ represent, as function of k, a probability distribution. A matrix that presents such a property is also called a *stochastic matrix*.

Many relations among distributions of different order can be put in matrix form. In particular, denoting

$$\pi_{jk}(n, n+1) = P(X(n) = j, X(n+1) = k),$$

we have

$$\pi_{jk}(n, n+1) = \pi_j(n)p_{jk}(n, n+1),$$

and summing up we get an expression of what is known as *Total Probability Theorem*:

$$\pi_k(n+1) = \sum_j \pi_j(n)\ p_{jk}(n, n+1). \tag{1.61}$$

By inspecting equation above we recognize the rule of the product of matrices, which allow us to rewrite (1.61) in matrix form as

$$\mathbf{\Pi}(n+1) = \mathbf{\Pi}(n)\mathbf{P}(n, n+1) \tag{1.62}$$

which is the fundamental formula of the evolution of the first-order distribution, since applied recursively starting from initial condions $\mathbf{\Pi}(0)$ provides

$$\mathbf{\Pi}(n) = \mathbf{\Pi}(0)\mathbf{P}(0, 1)\mathbf{P}(1, 2)\ldots\mathbf{P}(n-1, n). \tag{1.63}$$

We define matrix

$$\mathbf{P}(0, n) = \mathbf{P}(0, 1)\mathbf{P}(1, 2)\ldots\mathbf{P}(n-1, n) \tag{1.64}$$

whose elements

$$p_{jk}(0, n) = P(X(n) = k|X(0) = j),$$

represent the transition probabilities from time 0 to time $n+1$. Relation (1.63) can then be written as

$$\mathbf{\Pi}(n) = \mathbf{\Pi}(0)\mathbf{P}(0, n) \tag{1.65}$$

Definition (1.64) can be easily generalized to any pair of time instants n_1 and n_2 yielding $\mathbf{P}(n_1, n_2)$. The above relations can easily be extended providing the more general

$$\mathbf{\Pi}(n_2) = \mathbf{\Pi}(n_1)\mathbf{P}(n_1, n_2), \tag{1.66}$$

$$\mathbf{P}(n_1, n_3) = \mathbf{P}(n_1, n_2)\mathbf{P}(n_2, n_3), \tag{1.67}$$

the latter also known as Chapman-Kolmogorov equations.

1.5.3 Homogeneous Markov Chains

A very important case is when the mechanism of evolution is stationary, i.e., does not depend on n:

$$\mathbf{P}(n, n+1) = \mathbf{P}, \tag{1.68}$$

so that the entire Chain is completely described by the sole matrix $\mathbf{P}$. In this case, the MC is said *homogeneous*. Relations (1.62), (1.64) and (1.65) become respectively

$$\mathbf{\Pi}(n+1) = \mathbf{\Pi}(n)\mathbf{P} \tag{1.69}$$

$$\mathbf{P}(0, n) = \mathbf{P}^n, \tag{1.70}$$

$$\mathbf{\Pi}(n) = \mathbf{\Pi}(0)\mathbf{P}^n \tag{1.71}$$

The above show that all the evolution of a homogeneous MC is contained within $\mathbf{P}^n$. Furthermore, if at time 0 the state is known with probability one, which means that vector $\mathbf{\Pi}(0)$ is made of all zeros except for a one, say in position j, then $\mathbf{\Pi}(n)$ coincides with the j-th row of matrix $\mathbf{P}^n$.

Example (1.72)
Let assume that the state of a machine at each instant n is represented by a binary variable X where $X = 1$ means it is in operation and $X = 0$ it is not in operation (awaiting repair). Suppose that when working the machine may break with probability 0.2, while when in not working state it is repaired at each instant with probability 0.6, as a consequence of the availability of the repair resources.

The description above can be represented by matrix $\mathbf{P}(n, n+1)$ which is equal to

$$\mathbf{P}(n, n+1) = \begin{vmatrix} 0.4 & 0.6 \\ 0.2 & 0.8 \end{vmatrix}.$$

If we start at time $n = 0$ with the machine in operation, i.e.,

$$\pi_0(0) = 0, \qquad \pi_1(0) = 1,$$

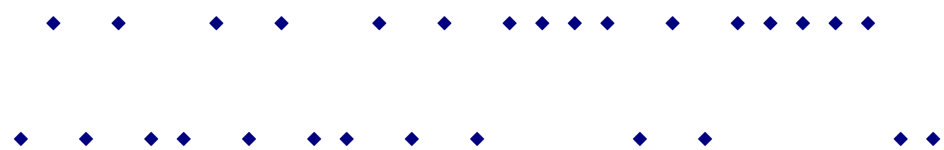


Figure 1.6: *Sample of a binary, purely-random, discrete-time process with first order distribution equal to 0.5*

Figure 1.7: *Sample of a binary, markovian, discrete-time process with "transition" probability equal to 0.2.*

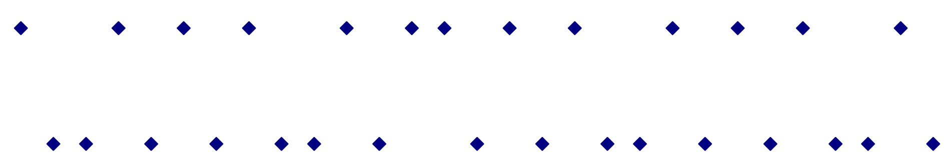


Figure 1.8: *Sample of a binary, markovian, discrete-time process with "transition" probability equal to 0.8.*

by (1.62) we can get $\pi_1(n)$, the probability that the machine is operating at time n. We have, for example,

$$\pi_1(1) = 0.8, \qquad \pi_1(2) = 0.76, \qquad \pi_1(3) = 0.752, \qquad \pi_1(4) = 0.7504. \quad \clubsuit$$

Example (1.73)

Consider transition matrices with all equal rows, such as

$$\mathbf{P}(n, n+1) = \begin{vmatrix} 0.8 & 0.2 \\ 0.8 & 0.2 \end{vmatrix}$$

Here any element of the matrix is such that $p_{jk} = p_k$, i.e., state k is reached with the same probability whichever the departing state j is. This means that no memory exists and the resulting process is purely random (in fact, there is no need to represent such process as markovian). Figure 1.6 shows sample of a binary, discrete-time chain with matrix

$$\mathbf{P}(n, n+1) = \begin{vmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{vmatrix} \clubsuit$$

Example (1.74)

Figures 1.7 and 1.8 show sample of a binary, markovian, discrete-time chain with

$$\mathbf{P}(n, n+1) = \begin{vmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{vmatrix}$$

in the former case and

$$\mathbf{P}(n, n+1) = \begin{vmatrix} 0.2 & 0.8 \\ 0.8 & 0.2 \end{vmatrix}$$

in the latter. If we start from vector

$$\mathbf{\Pi}(0) = [1/2; 1/2]$$

then we get in the two cases

$$\mathbf{\Pi}(1) = \mathbf{\Pi}(0)\mathbf{P}(n, n+1) = [1/2; 1/2]$$

$$\mathbf{\Pi}(n+1) = \mathbf{\Pi}(n)\mathbf{P}(n, n+1) = [1/2; 1/2]$$

The distribution of the first order is always $\pi_1(n) = 0.5$, as in the case of Figure 1.6, however, you can see that the state changes occur in different ways: with a strong "correlated memory" in the former case, i.e., trying to remain in that state and, and a strong "anti-correlated memory" in the latter case, i.e., trying to change it. Note that if the initial condition was different, distribution would be different. For example, if at time 0 the system is in 0, we have

$$\mathbf{\Pi}(0) = [0; 1]$$

and, for the second chain we have

$$\begin{aligned} \mathbf{\Pi}(1) &= \mathbf{\Pi}(0)\mathbf{P}(n, n+1) = [0.2; 0.8] \\ \mathbf{\Pi}(2) &= \mathbf{\Pi}(1)\mathbf{P}(n, n+1) = [0.68; 0.32] \qquad \clubsuit \\ \mathbf{\Pi}(3) &= \mathbf{\Pi}(2)\mathbf{P}(n, n+1) = [0.392; 0.608] \end{aligned}$$

Example *Random Walk* (1.75)
Let $X(n,\alpha) = X_n$ be a sequence of RVs' independent and equally distributed on values $(-1, 0, 1)$ where for each n we have

$$\begin{cases} P(X_n = 1) &= p \\ P(X_n = -1) &= q \\ P(X_n = 0) &= 1 - p - q \end{cases}$$

Process $Y(n,\alpha) = Y_n$ defined by:

$$\begin{cases} Y_0 = 0 \\ Y_{n+1} = Y_n + X_n \qquad n \geq 0 \end{cases} \tag{1.76}$$

is discrete-time process called *random walk* (in fact represents the position of a point that at each time moves at most one step up or down). This process is used to model different phenomena, such as the capital of a gambler that at every time bet a unit of money, that he wins or loses.

Figure 1.9 shows the description of the walk through the state diagram, which precisely shows the states and their transition probabilities in the case $p + q = 1$.

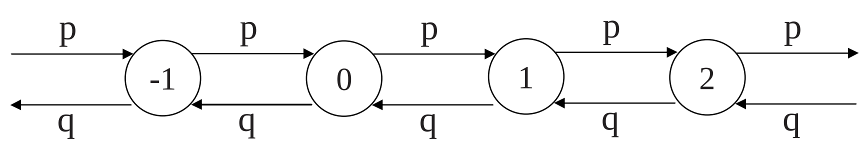


Figure 1.9: *State diagram of the random walk.*

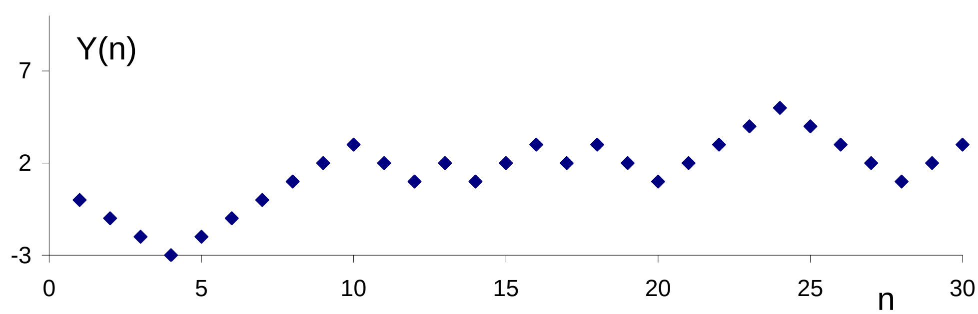


Figure 1.10: *Sample of a random walk with* $p = q = 0.5$

Figure 1.10 shows a sample of a random walk with $p = q = 0.5$. As can be seen memory is represented by the fact that the path (walk) can not much deviate from the previous position. A sample is easily obtained starting from its mathematical definition (1.76), once sequence $X(n)$ is given. The two cases in Example 1.74 also belong to some special random walk where the sum in $Y(n)$ is taken modulus 2, and $X(n)$ can only be either 0 or 1; in the first case we have $P(X(n) = 1) = 0.8$, while in the second $P(X(n) = 1) = 0.2$.

In a homogeneous Markov chain the exit from one state occurs independently from the past and with probability always equal to $\sum_{k \neq j} p_{jk} = 1 - p_{jj}$, exactly as Bernoulli arrivals. We therefore have

Property (1.77)
the continuous sojourn time Z_j in a state j, of a discrete-time homogeneous chain is a RV with Geometric distribution and mean value

$$E[Z_j] = \frac{1}{1 - p_{jj}} \tag{1.78}$$

1.5.4 Higher Order Markov Chains

Processes that present the following property

$$P(X_{n+1}|X_0, X_1, \ldots X_{n-1}, X_n) = P(X_{n+1}|X_{n-1}, X_n) \tag{1.79}$$

are called second-order Markov. Similarly, you can define Markov processes of higher orders.

The Markov processes of higher orders can be studied by transforming them into first-order Markov by changing the state variable to a new one that includes the memory of the original process. For example, a Markov process of the second order, $X(n)$, can be studied by defining process

$$Z(n) = [X(n); X(n-1)].$$

We can easily see (the proof is left to the reader) that $Z(n)$ obeys the (1.51) and is, therefore, Markov.

Example (1.80)
Let consider a random walk among states $0 \leq i \leq N$ with upward and downward probabilities equal to p and q respectively, with the modification that force the process not to leave the state the next step after entering the state; in fact, it can leave the state only in subsequent steps after the first (partial memory). The chain becomes Markov by suitably expanding the state space, whose diagram we must provide.

The knowledge of the present state is not enough to predict future behavior. In fact, we need to know whether the process is either just arrived into the state or if it was already there in the previous step. In other words we need to distinguish between the cases $X(n-1) = i, X(n) = i$ and $X(n-1) \neq i, X(n) = i$, which shows that also the past state $X(n-1)$ comes into play. To be able to tell the next step we must incorporate this "memory "in the state, which becomes now a two-dimensional one, namely (X, Y) where X represent the present state and Y, taking the value a and b, where $Y = a$ means that the present state is reached from another state, whereas and $Y = b$ means that the the previous state was the same. The state diagram of the new chain is shown in Figure 1.11. Denoted by ν_i the distribution of the original walk, and $\pi_{ix}, x = a, b$ the distribution of the new states, we have

$$\nu_i = \pi_{ia} + \pi_{ib} \qquad \clubsuit$$

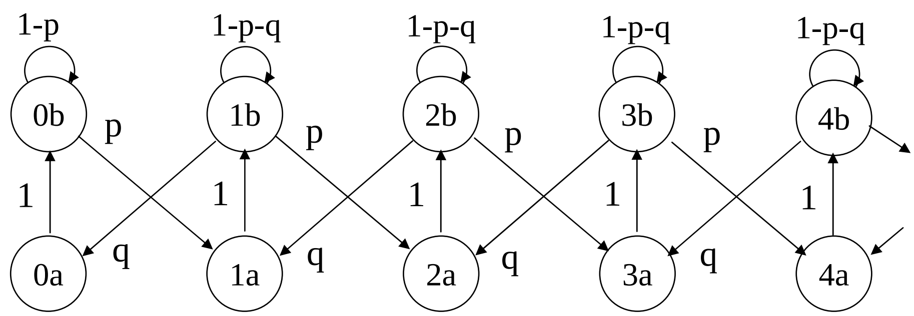


Figure 1.11: *State diagram of process* $(X;Y)$.

Example (1.81)
Figure 1.12a shows a sequence of letters drawn from a collection of the individual letters in the proportions used in the English language. It is also possible to measure the frequencies of letters that come after a given one, for all the letters in the first position (second order statistics). A compound sample of this type is in Figure 1.12b. Note, for example, that double spaces no longer appear because they do not exist in the statistical original. Similarly, using a model of the third order, the sample appears as shown in Figure 1.12c. Here you can note the appearance of words that make sense, such as -CAN- -OF- -TO-.

Example (1.82)
What is shown in the previous example can also be applied to words. Using the statistics of the recurring words in English texts, Shannon in 1948 obtained, with the purely random model, the sample shown in Figure 1.13a. Using the Markovian model of the first order, Shannon also obtained the sample shown in Figure 1.13b. This text has a greater coherence than the previous one, even if still provides little meaning.

Example (1.83)
Here we show an example similar to the previous one, with the statistics measured on the letters in a language we leave to the reader to guess (Abramson, 1963). The purly random sample is shown in Figure 1.14a. The second order sample is shown in Figure 1.14b, whereas the third order is in Figure 1.14c.

1.6 Continuous-time Chains

Here we considers chains in continuous time where transitions occur at times that are spaced by intervals with non-zero average, exactly as it happens with continuous-time arrivals. Therefore, transitions can be described by instantaneous rates, just as happens with continuous-time arrivals. For example, for state changes, in agreement with (1.23) we define the second-order transition probabilities as

$$p_{jk}(t, t+\Delta t) = q_{jk}(t)\Delta t + \omega(\Delta t), \qquad j \neq k. \tag{1.84}$$

AI-NGAE–ITF-NNR-ASAEV-OIE-BAINTHA-HYROO-POER-SETRYGAIETRWCO–EHDUARU-EU-C-FT-NSREM-DIY-EESE–F-O-SRIS-R–UNNASHOR

a

URTESHETHINH-AD-E-AT-FOULE-ITHALIORT-WACT-D-STE-MINSTAN-OLINS-TWID-OULY-TE-THIGHE-CO-YS-TH-HR-UPAVIDE-PAD-CTAVED

b

IANKS-CAN-OU-ANG-RLER-THATTED-OF-TO-SHOR-OF-TO-HAVEMEM-A-I-MAND-AND-BUT-WHISSITABLY-THERVEREER-EIGHTS-TAKILLIS-TA

c

Figure 1.12: *Approximations of the first, second and third order to the english language using a markov model on the letters of the english alphabet.*

REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT NATURAL HERE HE THE A IN CAME THE TO OF TO EXPERT GRAY COME TO FURNISHES THE LINE MESSAGE HAD BE THESE

a

THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED

b

Figure 1.13: *Approximations of the first and second to the english language using a markov model on the words.*

SETIOSTT-NINN-TUEEHHIUTIAUE-N-IREAISRI-
MINRNEMOSEPIN-MAIPSAC-SES-LN-ANEIISUNTINU-AR-TM-
UMOECNU-RIREIAL-AEFIITP

a

CT-QUVENINLUM-UA-QUREO-ABIT-SAT-FIUMAGE-ICAM-
MESTAM-M-QUM-QUTAT-PAM-NONDQUM-O-M-FIT-NISERIST-
E-L-ONO-IHOSEROCO

b

ET-LIGERCUM-SITECI-LIBEMUS-ACERELEN-TE-
VICAESCERUM-PE-NON-SUM-MINUS-UTERNE-UT-IN-ARION-
POPOMIN-SE-INQUENQUE-IRA

c

Figure 1.14: *Approximations of the first and second to the .... language.*

Since the sum of the rows of matrix $\mathbf{P}$ must be one, we also have

$$p_{jj}(t,t+\Delta t) = 1 - \sum_{k\neq j} p_{jk}(t,t+\Delta t) = 1 - \sum_{k\neq j} q_{jk}(t)\Delta t. \tag{1.85}$$

The result clearly shows that $p_{jj}(t,t+\Delta t) \to 1$ as $\Delta t \to 0$, in agreement with the fact that the transition out of state j occurs (with respect to the many $\Delta t's$) "rarely."

1.6.1 Continuous-time Markov Chains

Like discrete-time, continuous-time MC are completely described by second-order transition probabilities. Their relationship can be formally described as in the discrete-time case but changing n into t. For example, the analogous of (1.66) is

$$\mathbf{\Pi}(t_2) = \mathbf{\Pi}(t_1)\mathbf{P}(t_1,t_2). \tag{1.86}$$

Similarly, when we refer to the continuous-time analogous of a step, we must consider the interval Δt and (1.62) is rewritten as

$$\mathbf{\Pi}(t+\Delta t) = \mathbf{\Pi}(t)\mathbf{P}(t,t+\Delta t). \tag{1.87}$$

The equation (1.87) is correct only for $\Delta t \to 0$, and can not be used to get exact formulae. Instead, it can be re-written in a more correct way using derivatives. In fact, starting from

$$\frac{d\mathbf{\Pi}(t)}{dt} = \lim_{\Delta t\to 0} \frac{\mathbf{\Pi}(t+\Delta t) - \mathbf{\Pi}(t)}{\Delta t}, \tag{1.88}$$

and using (1.87) we have

$$\frac{\mathbf{\Pi}(t+\Delta t)-\mathbf{\Pi}(t)}{\Delta t}=\frac{\mathbf{\Pi}(t)\mathbf{P}(t,t+\Delta t)-\mathbf{\Pi}(t)\mathbf{I}}{\Delta t} \tag{1.89}$$

where $\mathbf{I}$ represents the identity matrix. Taking the limit we have

$$\frac{d\mathbf{\Pi}(t)}{dt}=\mathbf{\Pi}(t)\lim_{\Delta t\to 0}\frac{\mathbf{P}(t,t+\Delta t)-\mathbf{I}}{\Delta t}$$

and denoting

$$\mathbf{Q}(t)\triangleq\lim_{\Delta t\to 0}\frac{\mathbf{P}(t,t+\Delta t)-\mathbf{I}}{\Delta t} \tag{1.90}$$

we finally get the correct expression for the evolution equation in continuous time

$$\frac{d\mathbf{\Pi}(t)}{dt}=\mathbf{\Pi}(t)\mathbf{Q}(t). \tag{1.91}$$

where the entire description of the chain is now contained in the *transition-rates matrix* $\mathbf{Q}(t)$.

Relation (1.91), solved with some initial condition $\mathbf{\Pi}(0)$, allows to get the first order distribution at time t.

We see that matrix $\mathbf{Q}(t)$ is related to matrix $\mathbf{P}(t,t+\Delta t)$ by (1.90), whose elements are (see (1.84) and (1.85)):

$$\begin{aligned}
q_{jk}(t) &= \lim_{\Delta t\to 0}\frac{p_{jk}(t,t+\Delta t)}{\Delta t} && j\neq k\\
q_{jj}(t) &= \lim_{\Delta t\to 0}\frac{p_{jj}(t,t+\Delta t)-1}{\Delta t} = -\sum_{k\neq j} q_{jk}(t)
\end{aligned} \tag{1.92}$$

Note that, unlike matrix $\mathbf{P}$, now matrix $\mathbf{Q}$, has rows that add up to zero.

Relation (1.91), solved with some initial condition $\mathbf{\Pi}(0)$, allows to get the first order distribution at time t. However, analytical solutions are possible for rather simple cases only, such as the one shown in Example 1.93.

Example *Binary Chain* (1.93)
Let consider a binary chain $(0;1)$ where the intensity matrix does not depend on time. We have

$$q_{01}(t)=\lambda, \qquad q_{10}(t)=\mu$$

System (1.91) is written as

$$\frac{d\pi_0(t)}{dt} = +\pi_1(t)\mu - \pi_0(t)\lambda$$

$$\frac{d\pi_1(t)}{dt} = -\pi_1(t)\mu + \pi_0(t)\lambda$$

Since it is $\pi_o(t)+\pi_1(t) = 1$, the system above translates into the following linear differential equation

$$\frac{d\pi_1(t)}{dt} + (\lambda + \mu)\pi_1(t) = \lambda$$

whose solution is

$$\pi_1(t) = \frac{\lambda}{\lambda + \mu}[1 - e^{-(\lambda+\mu)t}] + \pi_1(0)e^{-(\lambda+\mu)t} \tag{1.94}$$

where $\pi_1(0)$ represents the initial condition. We also have

$$\pi_0(t) = \frac{\mu}{\lambda + \mu}[1 - e^{-(\lambda+\mu)t}] + \pi_0(0)e^{-(\lambda+\mu)t} \clubsuit \tag{1.95}$$

Example *Time-Continuous Random Walk* (1.96)
is the chain of Example 1.75 where the transition probabilities p and q are replaced by $\lambda\Delta t$ and $q = \mu\Delta t$. The analytical solution is quite involved and is omitted.

Example *Birth and Death (B&D) Process* (1.97)
is the generalization of a continuous-time random walk with transition rates that depend on the initial state in the following way $p_i = \lambda_i\Delta t$ e $q_i = \mu_i\Delta t$.$\clubsuit$

Example *Pure Birth Process, Poisson Counting Process* (1.98)
is a $B\&D$ process with $\lambda_i = \lambda$ and $\mu_i = 0$, and it represents the number of births in between times 0 and t. System (1.91) writes scrive

$$\frac{d\pi_0(t)}{dt} = -\pi_0(t)\lambda$$

$$\frac{d\pi_i(t)}{dt} = \pi_{i-1}(t)\lambda - \pi_i(t)\lambda \qquad i = 1, 2, \ldots$$

The equations above have a recursive solution simple enough. With the initial conditions

$$\pi_0(0) = 1$$

$$\pi_i(0) = 0 \qquad i = 1, 2, \ldots$$

we get

$$\pi_i(t) = \frac{(\lambda t)^i}{i!}e^{-\lambda t}. \tag{1.99}$$

The result is obvious, since the conditions of this process correspond to the axioms of Poisson arrivals.

1.6.2 Homogeneous continuous-time Markov Chains

Time-continuous and homogeneous Markov chains are those where matrix $\mathbf{Q}(t)$ does not depend on t, and equation (1.91) becomes

$$\frac{d\mathbf{\Pi}(t)}{dt} = \mathbf{\Pi}(t)\mathbf{Q}. \tag{1.100}$$

If we refer to the occurrence of a transition from a specific state j to a specific state $k =\neq j$, then we see that those occurrences can be seen exactly as Poisson arrivals of rate q_{jk}. In particular, the the residence time Z_{jk}, i.e., the time between the entrance in state j and the transition toward state k, represents the distance between two consecutive arrivals. Then we have

Property (1.101)
In a homogeneous continuous-time chain the residence time Z_{jk} is a negative exponential RV with mean value

$$E[Z_{jk}] = \frac{1}{q_{jk}} \tag{1.102}$$

In the same way, the exits from state j can be seen as a superposition of transitions (arrivals) jk, and can be seen exactly as Poisson arrivals of rate $-q_{jj} - \sum_{k\neq j} q_{jk}$. In particular, the residence time Z_j, i.e., the time between the entrance in state j and the transition toward any other state k, represents the distance between two consecutive such arrivals. Then we have

Property (1.103)
In a homogeneous continuous-time chain the residence time Z_j is a negative exponential RV with mean value

$$E[Z_j] = \frac{1}{-q_{jj}} \tag{1.104}$$

Discrete time Chains can be simulated step by step determining at each time the state according to a distribution conditional to the entire past, or, for MC, conditional to the last step. Such procedure can not be retained with continuous-time chains; in fact assuming discrete steps Δt can not provide an exact simulation. However, Property (1.102) suggests how to simulate a homogeneous continuous-time a chain; in fact, once entered state j, it suffices to simulates all the negative exponential RVs representing all residence times Z_{jk}. The transition occurs at the transition time that comes first and the new reached state is the one related to such time. The procedure is the repeated with the new state. This procedure is expensive if states k are very many. Alternatively, we can simulate the negative exponential RV Z_j at which a transition occur. Since transition to state k occurs with probability $q_{jk}/(-q_{jj})$. One additional experiment provides such state.

1.7 Stationary Processes

If the description of a process, for example in terms of pdf's of its RVs, does not change while shifting the origin of the axis, the process is said *stationary*. If this is the case the distribution

$\mathbf{\Pi}(k, n_2 + k, \ldots, n_r + k)$ does not depend on k, and we have

$$\mathbf{\Pi}(n) = \mathbf{\Pi} \qquad \forall \;\; n \tag{1.105}$$

$$\mathbf{\Pi}(m, n) = \mathbf{\Pi}(n - m) \qquad \forall \;\; n, m. \tag{1.106}$$

$$\mathbf{P}(m, n) = \mathbf{P}(n - m) \qquad \forall \;\; n, m. \tag{1.107}$$

and so on. Note that, in these conditions, relation (1.62) becomes

$$\mathbf{\Pi} = \mathbf{\Pi}\mathbf{P}, \tag{1.108}$$

and we also have

$$\mathbf{\Pi} = \mathbf{\Pi}\mathbf{P}^r \qquad \forall \;\; r. \tag{1.109}$$

A distribution that obeys to (1.108) is called "stationary".

Example (1.110)
Verify that the following binary chain with transition matrix

$$\mathbf{P} = \begin{vmatrix} 0.2 & 0.8 \\ 0.8 & 0.2 \end{vmatrix}$$

has vector $[0.5; 0.5]$ *as stationary distribution.*♣

The process is then called *asymptotically stationary* when, for every r, the following limit

$$\lim_{n\to\infty} \mathbf{\Pi}(n_1 + n, n_2 + n, \ldots, n_r + n) \tag{1.111}$$

turns out to be independent of n. Therefore, for the first order we have

$$\lim_{n\to\infty} \mathbf{\Pi}(n) = \mathbf{\Pi} \qquad \forall \;\; n \tag{1.112}$$

Example (1.113)
With reference to the chain of Example 1.110, find the $\mathbf{\Pi}(n)$ *for* $n = 1, 3, 5, 10, 20$*, starting from the initial condition* $\mathbf{\Pi}(0) = [1; 0]$.

Using

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

we get $\mathbf{\Pi}(1) = [0.2; 0.8]$, $\mathbf{\Pi}(3) = [0.3920; 0.6080]$, $\mathbf{\Pi}(5) = [0.4611; 0.5389]$, $\mathbf{\Pi}(10) = [0.5030; 0.4970]$, $\mathbf{\Pi}(20) = [0.5000; 0.5000]$. ♣

Example *Time continuous binary chain* (1.114)
With reference to Example 1.93 we see that such chain is asymptotically stationary whatever the initial condition are. In fact, we have

$$\lim_{t\to\infty} \pi_1(t) = \frac{\lambda}{\lambda + \mu}, \tag{1.115}$$

$$\lim_{t\to\infty} \pi_0(t) = \frac{\mu}{\lambda+\mu}. \quad \clubsuit \tag{1.116}$$

Property (1.117)
the reversed process of a stationary MC with transition probabilities p_{ij} and first order distribution π_i is an MC with homogeneous transition probabilities

$$p'_{ij} = p_{ji}\frac{\pi_j}{\pi_i}. \tag{1.118}$$

This comes from equation (1.118) (Bayes), here rewritten by expliciting states, which in stationary conditions becomes

$$P(X_{n-1}=j|X_n=i) = P(X_n=i|X_{n-1}=j)\frac{P(X_{n-1}=j)}{P(X_n=i)} = p_{ji}\frac{\pi_j}{\pi_i}, \qquad \forall n.\clubsuit$$

Again denoting by prime the distributions of the reversed process, if it happens that,

$$\pi'_{ij} = \pi_{ij}, \qquad \forall i,j, \tag{1.119}$$

the second order distribution does not change upon a reversal of time axis, which means that the reversed process is probabilistically indistinguishable from the original one, and is called *reversible*.

Since it is $\pi'_i = \pi_i$, relation (1.119) can be written

$$p'_{ij} = p_{ij},$$

and using (1.118) we are able to prove the following

Property (1.120)
a stationary MC is reversible if and only if we have:

$$\pi_j p_{ji} = \pi_i p_{ij} \qquad \forall i,j. \tag{1.121}$$

Referring to Example 1.110, the chain is easily proved reversible.

Example (1.122)
The chain with transition matrix

$$\mathbf{P} = \begin{vmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{vmatrix}$$

has a stationary distribution equal to $[1/3\ 1/3\ 1/3]$. *Relation (1.119) is never satisfied and the chain is non-reversible.* $\clubsuit$

A condition similar to (1.119), using intensities q_{ij}, holds for the reversibility of continuous-time MCs.

1.8 Homogeneous and Irreducible Finite Markov Chains

Homogeneous Markov Chains evolve according to the law given by (1.71) which we re-write below:

$$\mathbf{\Pi}(n) = \mathbf{\Pi}(0)\mathbf{P}^n \tag{1.123}$$

The evolution of finite chains occurs according to three types of patterns, which are illustrated by the three examples below. In particular, we are interested in finding, if exists an asymptotic behavior, i.e.,

$$\lim_{n\to\infty} \mathbf{\Pi}(n) = \mathbf{\Pi}(0) \lim_{n\to\infty} \mathbf{P}^n = \mathbf{\Pi}^* \tag{1.124}$$

$$\lim_{n\to\infty} \mathbf{P}^n = \mathbf{P}^* \tag{1.125}$$

Should limit (1.125) exists, also (1.124) exists and we have

$$\mathbf{\Pi}^* = \mathbf{P}^*\mathbf{\Pi}(0), \tag{1.126}$$

where $\mathbf{\Pi}^*$ is called *asymptotic distribution* and $\mathbf{\Pi}(0)$ the *asymptotic matrix.*

The above shows that the asymptotic behavior of the chain depends on $\mathbf{P}^*$, which can be of three different types, namely:

- all rows of $\mathbf{P}^*$ are equals, which means that the starting state, (the index of the row) does not matter. In this case $\mathbf{\Pi}^*$ is always the same whichever $\mathbf{\Pi}(0)$;

- some rows of $\mathbf{P}^*$ differs, which means that the starting state does matter. In this case $\mathbf{\Pi}^*$ depends on $\mathbf{\Pi}(0)$;

- $\mathbf{P}^*$ does not exists, because it keeps changing at any n.

Example (The ruin problem) (1.127)
Two players bet against each other, having probability $1/2$ *to win. Each time they bet a unit and the game ends when one of the two remains without capital. If the first player starts with capital* c*,* $c = 0, 1, 2, 3$ *and the second player with capital* $3 - c$*. When a player reaches capital zero he can no longer bet and its state reains in zero forever. The evolution of the capital of the first player is a MC whose transition matrix has size* 4×4*. Write the transition matrix in* $1, 2, 10, 20$ *steps.*

We have

$$\mathbf{P} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 1 \end{vmatrix} \qquad \mathbf{P}^2 = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0.5 & 0.25 & 0 & 0.25 \\ 0.25 & 0 & 0.25 & 0.5 \\ 0 & 0 & 0 & 1 \end{vmatrix}$$
$$\mathbf{P}^{10} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0.666 & 0.001 & 0 & 0.333 \\ 0.333 & 0 & 0.001 & 0.666 \\ 0 & 0 & 0 & 1 \end{vmatrix} \qquad \mathbf{P}^{20} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0.667 & 0 & 0 & 0.333 \\ 0.333 & 0 & 0 & 0.6667 \\ 0 & 0 & 0 & 1 \end{vmatrix} \tag{1.128}$$

We note that $\mathbf{P}^n$ after some steps seems to settle to an asymptotic matrix whose rows all different, each one of them, we recall, represents the asymptotic distribution when departing from the state represented by the row.

Example (The urn problem) (1.129)
Two urns contain each N balls, N of which are blue and the others red, subdivided at random between the urns. At each instant, draw a ball from both urns and, with probability $1/2$ exchange urn. The number of red balls in the first urn, $X(n)$ is Markov chain. Assuming $N = 2$, write the transition matrix in $1, 2, 10, 20$ steps.

We have

$$\mathbf{P} = \begin{vmatrix} 0.5 & 0.5 & 0 \\ 0.125 & 0.75 & 0.125 \\ 0 & 0.5 & 0.5 \end{vmatrix} \qquad \mathbf{P}^2 = \begin{vmatrix} 0.3125 & 0.625 & 0.0625 \\ 0.15625 & 0.6875 & 0.15625 \\ 0.0625 & 0.625 & 0.3125 \end{vmatrix}$$
$$\mathbf{P}^{10} = \begin{vmatrix} 0.1672 & 0.6667 & 0.1662 \\ 0.1667 & 0.6667 & 0.1667 \\ 0.1662 & 0.6667 & 0.1672 \end{vmatrix} \qquad \mathbf{P}^{20} = \begin{vmatrix} 0.1667 & 0.6667 & 0.1667 \\ 0.1667 & 0.6667 & 0.1667 \\ 0.1667 & 0.6667 & 0.1667 \end{vmatrix} \tag{1.130}$$

Here we see that $\mathbf{P}^n$ after some steps seems to settle to an asymptotic matrix $\mathbf{P}^*$ whose rows all equal. This means that, whichever the initial condition are, asymptotically we reach the same distribution.

Example (The bounded random walk) (1.131)
Let consider a random walk with reflecting barriers at 0 and 3, where the steps are always ± 1, with probability 0.5, except for the boundaries where the step is always backward. Write the transition matrix in $1, 2, 20, 21$ steps.

We have

$$\mathbf{P} = \begin{vmatrix} 0 & 1 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 1 & 0 \end{vmatrix} \qquad \mathbf{P}^2 = \begin{vmatrix} 0.5 & 0 & 0.5 & 0 \\ 0 & 0.75 & 0 & 0.25 \\ 0.25 & 0 & 0.75 & 0 \\ 0 & 0.5 & 0 & 0.5 \end{vmatrix}$$
$$\mathbf{P}^{20} = \begin{vmatrix} 0.3333 & 0 & 0.6667 & 0 \\ 0 & 0.6667 & 0 & 0.3333 \\ 0.3333 & 0 & 0.6667 & 0 \\ 0 & 0.6667 & 0 & 0.3333 \end{vmatrix} \qquad \mathbf{P}^{21} = \begin{vmatrix} 0 & 0.6667 & 0 & 0.3333 \\ 0.3333 & 0 & 0.6667 & 0 \\ 0 & 0.6667 & 0 & 0.3333 \\ 0.3333 & 0 & 0.6667 & 0 \end{vmatrix} \tag{1.132}$$

Here $\mathbf{P}^n$ does not reach a limit; in fact, after a while, it changes alternately assuming the shapes we see in $n = 20$ and $n = 21$. ♣

It is left to the reader to verify that if $p = q < 0.5$, i.e., the chain is allowed to remain in a state, then the rows of $\mathbf{P}^n$ tend to the same limit and we have the same case as the Urn problem. The difference with the last two examples is that we still have a limiting behavior, which is periodic, but not an asymptotic distribution.

Looking for the causes that make the evolution independent of initial conditions, we see that this is due to the fact that starting from any state, any other state can be reached within a finite amount of time in the average. If this happen, then initial conditions must not influence asymptotic; in fact, if we start from state j, after a finite amount of time we reach state k; then the asymptotic is the same as we started from k. What just described can not happen if some state can not be reached by some other, as it happens in the ruin problem. The property above can be ascertained by introducing irreducible chains.

State j is said to communicate with k if we have $p_{jk}(n) > 0$ for some n. A set of states is said *irreducible* if and only if is composed by states that communicate with any other within the set and with no other outside the set. Then we have the following

Property (1.133)
In a finite irreducible Markov chain each state can be reached, in the end, from any other state with probability one and after a finite amount of time in the average.

This property is intuitive, but its proof requires some detailed passages that are not exposed here; it can be proved by observing that the Markov property, especially the one in (1.101) assures that we exit any state j within a finite amount of time in the average, and by the fact that we have a finite number of states.

The same Property above explains why

Property (1.134)
in an irreducible chain we can have either an asymptotic behavior or an asymptotically periodic one.

In fact, if the return to any state k occurs within a finite amount of time, and N is the number of states, we can only have at most N different asymptotically periodic behaviors, one for each departing state N, and at least only one behavior, the asymptotic one.

If in the limit the initial conditions do no matter then, then the limiting matrix $\mathbf{P}^*$ has by definition all equal rows and each row coincides with the limiting distribution $\mathbf{\Pi}^*$. Hence we have proved the following

Property (1.135)
If a homogeneous Markov chain is irreducible and non-periodic the asymptotic distribution $\mathbf{\Pi}^*$ *always exists, is unique and independent from initial conditions. This also means that all rows of* $\mathbf{P}^*$ *are equal to* $\mathbf{\Pi}^*$*, whatever* $\mathbf{\Pi}(0)$ *is.*

The irreducible property can be often easily ascertained looking at the state diagram.

Now the problem arises to find the asymptotic distribution without having to evaluate the limit of the distribution, since this is quite often impractical.

1.8.1 Balance equations

Referring to (1.87) we also have:

$$\lim_{n\to\infty} \mathbf{\Pi}(n+1) = (\lim_{n\to\infty} \mathbf{\Pi}(n))\mathbf{P}, \tag{1.136}$$

which shows that

Property (1.137)
If an asymptotic distribution exists, then it is also a stationary distribution.

We call stationary, or equilibrium distribution, any distribution $\mathbf{\Pi}$ that obey the relation

$$\mathbf{\Pi} = \mathbf{\Pi}\mathbf{P} \tag{1.138}$$

We have,

Property (1.139)
if the asymptotic distribution is unique also the equilibrium distribution is unique and coincides with the former.

In fact, if it were not so, taking one of equilibrium distribution, say $\mathbf{\Pi}_l$, as one initial condition $\mathbf{\Pi}(0) = \mathbf{\Pi}_l$, we would have $\mathbf{\Pi}(1) = \mathbf{\Pi}_l$, $\mathbf{\Pi}(2) = \mathbf{\Pi}_l$, $\ldots \mathbf{\Pi}(n) = \mathbf{\Pi}_l$, that is we should have many asymptotic distributions, which is a contradiction.

Therefore, it may happen that, when dealing with irreducible chains, the two terms, asymptotic and stationary, are used interchangeably.

Since in an irreducible chain all states are visited, there can not be a state whose asymptotic probability is zero. Otherwise all the other states should also have zero asymptotic probability. This can be ascertained by the relation (1.138), that in scalar form is

$$\pi_k = \sum_j \pi_j p_{jk}. \tag{1.140}$$

In order to have $\pi_k = 0$ we must have $\pi_j = 0$ for all j such that $p_{jk} > 0$. Repeating this argument we should have $\pi_j = 0$ for all j. Hence:

Property (1.141)
all the elements of the asymptotic distribution of a homogeneous irreducible and finite MC are greater than zero.

Relation (1.138) suggests to find the stationary (and asymptotic) distribution solving the system of equations

$$\mathbf{X} = \mathbf{X}\mathbf{P} \tag{1.142}$$

The system of equations (1.142) is called the equilibrium conditions, or balance equations system. It clearly admits the stationary solution $\mathbf{\Pi}$, or more if the chain is not irreducible. However, even with irreducible chains, it also admits as solution the identically zero vector, and if admits a non-null solution $\mathbf{X}$ it also admits as solution the proportional sequence $k\mathbf{X}$. This multiplicity of solutions exists because the rows of the matrix $\mathbf{P}$ are not linearly independent since all add up to 1. In other words, if N is the number of states, we only have $N-1$ independent equations with N unknowns. We need an additional independent equation to get the unique distribution, and this is provided by the congruence relation

$$\sum_j \pi_j = 1, \tag{1.143}$$

that traduces into

$$\sum_j kX_j = 1$$

and provides

$$\pi_i = \frac{X_i}{\sum_j X_j} \tag{1.144}$$

This is the same as to substitute any of the scalar equations in (1.142) with (1.143). Therefore, we have proved the following

Theorem: (1.145)
In an homogeneous, irreducible and finite MC, the asymptotic distribution is obtained by solving the system of equations (1.142) *with* (1.144).

Example (1.146)
Find the asymptioc distribution of the binary chain with $p_{01} = p$ *and* $p_{10} = q$.

we have

$$\pi_0 = (1-p)\pi_0 + q\pi_1$$

$$\pi_0 + \pi_1 = 1$$

which yield

$$\pi_0 = \frac{q}{p+q}, \qquad \pi_1 = \frac{p}{p+q} \clubsuit$$

1.8.2 Equilibrium of probability fluxes

Condition (1.108),

$$\mathbf{\Pi} = \mathbf{\Pi P} \tag{1.147}$$

or, in scalar form,

$$\pi_k = \sum_j \pi_j p_{jk}, \quad \forall k, \tag{1.148}$$

suggests an interesting interpretation. The term

$$\pi_j p_{jk} = \pi_{jk}$$

is called *probability flux* out of state j toward state k. Then we call *cut* on the state space, a sub-division of the whole set $\mathcal{S}$ of states into two complementary sub-sets $\mathcal{A}$ e $\mathcal{B} = \mathcal{S} - \mathcal{A}$. In these conditions, (1.147) and (1.148) express the fact that

Theorem: (1.149)
Under stationary conditions, the total algebraic flux through any cut in state space is zero. (That is why the stationary distribution is also called the "equilibrium" distribution).

Proof
The equilibrium system of equations

$$\pi_k = \sum_{i \in \mathcal{S}} \pi_i \; p_{ik},$$

can be re-written as

$$\pi_k \sum_{i \in \mathcal{S}} p_{ki} = \sum_{i \in \mathcal{S}} \pi_i \; p_{ik} \tag{1.150}$$

where we have used the identity $\sum_{i \in \mathcal{S}} p_{ki} = 1$. If we sum both sides over $k \in \mathcal{A}$, we get

$$\sum_{k \in \mathcal{A}} \sum_{i \in \mathcal{S}} \pi_k p_{ki} = \sum_{k \in \mathcal{A}} \sum_{i \in \mathcal{S}} \pi_i \; p_{ik}.$$

and, erasing the common terms, we finally get:

$$\sum_{k \in \mathcal{A}} \sum_{i \in \mathcal{S}-\mathcal{A}} \pi_k p_{ki} = \sum_{k \in \mathcal{A}} \sum_{i \in \mathcal{S}-\mathcal{A}} \pi_i \; p_{ik} \tag{1.151}$$

That shows indeed the equilibrium of fluxes. ♣

Note that, if N is the number of states, once the balance of fluxes in and out $N-1$ states is imposed, then the balance of fluxes in and out the remaining state is granted. Which means that the N-th balance equation is already contained in the preceding $N-1$ ones, or, in other words, any relation in (1.148) can be obtained by a linear combination of the other ones. Of course, this is true also for other cuts, so that to write $N-1$ independent balance equations we need $N-1$ independent cuts.

Also note that the original equation (1.148) also represents the balance of fluxes. In fact, written in the form (1.150), it shows the equilibrium of fluxes in and out state k, i.e., across the cut that isolates state k.

The above Theorem shows that equations in (1.148) (for any k) can be replaced by balance equations across an equivalent number of independent cuts, with the advantage that balance equations at cleverly selected cuts can be simpler than the original ones, for examples involving less terms.

Theorem 1.149 is the analogue of Kirckoff Law for the electrical currents flowing into a node of an electrical network.

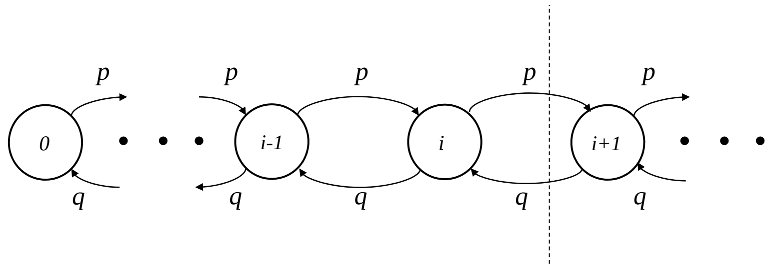


Figure 1.15: Tagli dello spazio degli stati relativo alla passeggiata casuale.

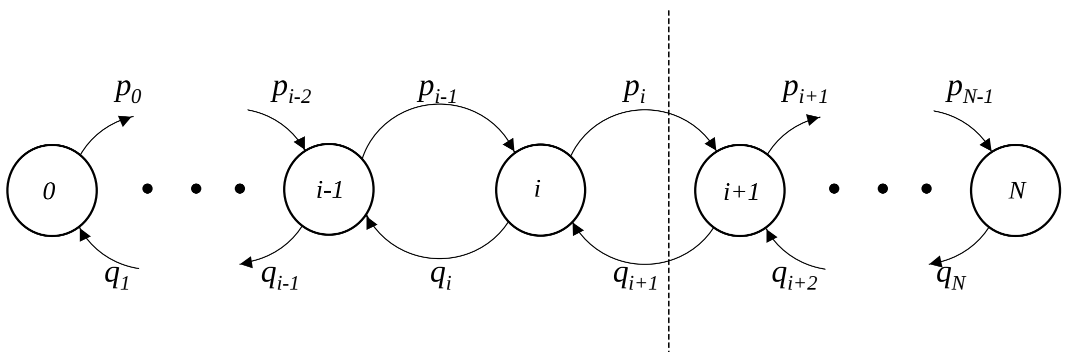


Figure 1.16: Transition diagram of the generalized random walk.

Example *Generalized random walk* (1.152)
The state transitions are shown in Figure1.16. We assume that the walk take places within limit states 0 and N. The transition mpatrix is

per $i = 1, 2, \dots, N-1$

$$\begin{aligned}
p_{i,i+1} &= p_i \\
p_{i,i-1} &= q_i \\
p_{i,i} &= 1 - p_i - q_i
\end{aligned}$$

furthermore:

$$\begin{aligned}
p_{0,1} &= p_0 \\
p_{0,0} &= 1 - p_0 \\
p_{N,N-1} &= q_N \\
p_{N,N} &= 1 - q_N
\end{aligned}$$

all the others being zero.

We impose the equilibrium of fluxes across cuts such as the one in Figure 1.16, and have

$$\pi_i p_i = \pi_{i+1} q_{i+1} \tag{1.153}$$

System (1.153) can be solved recursively as function of π_0, starting from

$$\pi_1 = \pi_0 \frac{p_0}{q_1}.$$

We get :

$$\pi_i = \pi_0 \frac{p_0\ p_1\ p_2 \cdots p_{i-1}}{q_1\ q_2 \cdots q_i} \qquad i = 1, \ldots, N \tag{1.154}$$

Probability π_0 is derived from the condition $\sum_k \pi_k = 1$, which provides

$$\pi_0 = \frac{1}{1 + \displaystyle\sum_{i=1}^{N} \frac{p_0 p_1 \cdots p_{i-1}}{q_1 q_2 \cdots q_i}}. \tag{1.155}$$

Example *Random Walk within boundaries* (1.156)
Can be derived by the preceding example setting $p_i = p$ and $q_i = q$. Relations (1.154) and (1.155) provide:

$$\begin{cases} \pi_i &= (p/q)^i \dfrac{1 - p/q}{1 - (p/q)^{N+1}}, \qquad i = 0, 1, \ldots, N, \qquad p \neq q \\ \\ \pi_i &= \dfrac{1}{N+1}, \qquad i = 0, 1, \ldots, N, \qquad p = q. \end{cases} \tag{1.157}$$

Distribution (1.157), is the *truncated geometric* distribution for $p \neq q$ and the uniform one for $p = q$, and is the unique asymptotic and stationary distribution of this walk.♣

Note that if $p + q = 1$ the walk shows periodicity. In fact, if we start from state k, you can not return to the same state k, or go to states $k + 2r$, in other than an even number of steps. So there can not exist an asymptotic distribution in the strict sense. The significance of the distribution found with the system of equations is clarified later on.

Referring to the Generalized Random Walk of Example 1.152, we have seen that cuts of the type shown in Figure 1.16 lead to balance equations such as the (1.153), which naturally satisfy the condition (1.119) for reversibility, and then the stationary Generalized Random Walk is a reversible process. We note that a distribution π_i satisfying (1.119), i.e. which balance probability fluxes between any couple of states, automatically satisfy the more general balance equations (1.148) that represent the condition for stationarity. This proves that the set of reversible MCs is a subset of the stationary ones.

Example (1.158)
We now find the asymptotic distribution of chain $X(n)$ of Example 1.80. We study chain $Y(n)$ whose state diagram is shown in Figure 1.17.

To simplify the solution, balance equations must be chosen in a suitable way. Balance equation at nodes ia show that all flux entering the node goes into node ib. This suggest a shortcut. If we

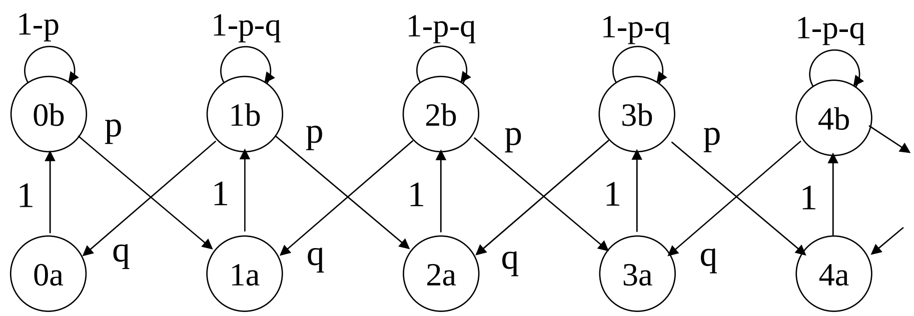


Figure 1.17: *State diagram of process* $(X;Y)$.

drop states ia and set the corresponding incoming fluxes into states ib, now called ic, then, nodes ic show a diagram that is exactly the one of random walk. Furthermore, if fluxes balance in Figure 1.17, they also balance in the reduced state space ic of the random walk . This means that the equlibrium fluxes of both systems are the same, but a proportionality constant. Since we have

$$\pi_{ic} = \pi_{0c}\left(\frac{p}{q}\right)^i, \qquad i \geq 0.$$

We also have

$$\pi_{ib} = k\left(\frac{p}{q}\right)^i, \qquad i \geq 0.$$

The equilibrium at nodes ia is

$$\pi_{ia} = \pi_{i-1b}p + \pi_{i+1b}q, \qquad i \geq 1,$$

$$\pi_{0a} = \pi_{1b}q = kp,$$

or

$$\pi_{ia} = k\left(\frac{p}{q}\right)^{i-1} p + k\left(\frac{p}{q}\right)^{i+1} q = k\left(\frac{p}{q}\right)^i (q+p), \qquad i \geq 1,$$

Now we can sum over all π_{ia} and π_{ib}, to get k. However, the distribution of the original chain $X(n)$ is

$$\pi_i = \pi_{ia} + \pi_{ib} = k\left(\frac{p}{q}\right)^i (q+p+1), \qquad i \geq 1,$$

$$\pi_0 = k(p+1).$$

and

$$k(p+1) + \sum_{i=1}^{\infty} k\left(\frac{p}{q}\right)^i (q+p+1) = 1$$

which provides k and the whole distributions. ♣

Example *Doubly stochastic matrices* (1.159)
If the columns of a stochastic matrix sum to one the matrix is said doubly stochastic. For such matrices a solution of balance equations

$$x_j = \sum_k x_k p_{kj}$$

is $x_j = 1, \ \forall j$, because $\sum_k p_{kj} = 1$. Therefore, the solution is uniform, equal to the reverse of the number of states. This is the case of Random Walk with $p = q$.

Example (1.160)
A random Walk on a circle with N states is such that at each time the state is advanced by one (or from $N-1$ to 0) with probability p; otherwise, with probability $1-p$ the state is not changed.

It is easily recognized that the transition matrix is doubly stochastic. Hence his distribution is uniform.

1.8.3 Balance equations with continuous-time MCs

Balance equations (1.138) for discrete-time chains originates from the limit (1.136). Similarly, (1.100), applied to the asymptotic distribution $\mathbf{\Pi}$, which can not depend on time, becomes

$$\mathbf{\Pi Q} = 0. \tag{1.161}$$

In an alternative way, the asymptotic distribution $\mathbf{\Pi}$ also satisfy the following

$$\mathbf{\Pi} = \mathbf{\Pi P}(\Delta t), \tag{1.162}$$

which can be re-conducted to (1.161) in the following way: using the scalar form

$$\pi_k = \sum_j \pi_j \ p_{jk}(\Delta t) \tag{1.163}$$

and replacing with transition rates, the above relation becomes

$$\pi_k = \sum_{j \neq k} \pi_j \ q_{jk} \Delta t + \pi_k p_{kk}(\Delta t) = \sum_{j \neq k} \pi_j \ q_{jk} \Delta t + \pi_k (1 - \sum_{i \neq k} q_{ki} \Delta t).$$

and, simplifying

$$0 = \sum_{j \neq k} \pi_j \ q_{jk} - \pi_k \sum_{i \neq k} q_{ki}. \tag{1.164}$$

Using relation

$$q_{kk} = -\sum_{i \neq k} q_{ki},$$

terms kk may be taken into the summation and the whole RHS of (1.164) becomes a matrix product that can be written as

$$0 = \mathbf{\Pi Q} \tag{1.165}$$

Equation (1.165) is, therefore, completely equivalent to (1.162), but does not make use of terms Δt and can be more correctly used as equilibrium equation. Being it derived from (1.165) shows that it still expresses the balance of the probability fluxes, which now can be more correctly defined as $\pi_k q_{jk}$, having dropped the term Δt, that appears in all fluxes and that does not alter the balance.

What said above allows us to state that:

Theorem: (1.166)
the stationary distribution of a continuous-time MC is attained exactly as with a discrete-time chain where transition rates are used in place of transition probabilities.

Example (1.167)
The above shows that the solution in Example 1.187 for the generalized walk also provides the solution in the continuous time as

$$\begin{cases} \pi_i &= (\lambda/\mu)^i \dfrac{1-\lambda/\mu}{1-(\lambda/\mu)^{N+1}}, \qquad i=0,1,\ldots,N, \qquad \lambda \neq \mu \\ \\ \pi_i &= \dfrac{1}{N+1}, \qquad i=0,1,\ldots,N, \qquad \lambda = \mu. \end{cases} \tag{1.168}$$

We can also check that the stationary solution of the binary chain above coincides with the asymptotic solution found in (1.115) and (1.116).

Example *Ehrenfest diffusion model* (1.169)
A box is composed by two parts connected by a hole. The box contains M molecules, and $N(t)$, $0 \leq N(t) \leq M$ is number of molecules in the first part of the box. *Each of them*, after a negative exponential time of rate η, independent for each molecule, goes to the other side of the box through the hole; both ways are possible.

The state diagram is that of a Birth and Death process with

$$\lambda_i = (M-i)\eta, \qquad \mu_i = i\eta.$$

The distribution is derived from the general solution of the Birth and Death process:

$$\pi_i = \pi_0 \frac{M(M-1)\ldots(M-i+1)}{i!} = \binom{M}{i} 2^{-M},$$

which is the Binomial distribution with $p = 1/2$.

1.9 Non-Irreducible Markov Chains with finite state space

If the state space is not irreducible, then it can be proven that it is composed by two or more irreducible subsets of states, that are absorbing subsets, and a *transient* subset. In fact we have the following

Property (1.170)
Starting from any transient state, within a finite amount of time in the average the chain reaches either of the irreducible subsets and never leaves.

After reaching a specific irreducible subset $\mathcal{A}$ the asymptotic distribution becomes the one relative to subset $\mathcal{A}$. The probability at which $\mathcal{A}$ is reached, $P_{\mathcal{A}}$, depends on the departing state. Therefore, in the non-periodic case, the asymptotic distribution starting from a transient state can be attained by the Total Probability Theorem.

Example (1.171)
Find the asymptotic distribution of the discrete-time MC whose state diagram shown in figure 1.18, when starting in states $1, 2, 3$.

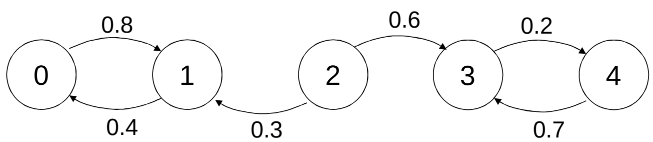


Figure 1.18:

The chain is non-irreducible with two irreducible subsets, namely $\{0, 1\}$ and $\{3, 4\}$, while state 2 is transient. According to what we said above, when starting in state 1 and 3 the asymptotic distribution is respectively

$$\mathbf{\Pi} = [1/3 \;\; 2/3 \;\; 0 \;\; 0 \;\; 0]; \qquad \mathbf{\Pi} = [0 \;\; 0 \;\; 0 \;\; 7/9 \;\; 2/9].$$

When starting in state 2, the chain is absorbed in $\{0, 1\}$ with probability $3/9$ and in $\{3, 4\}$ with probability $6/9$. hence the asymptotic distribution is

$$\mathbf{\Pi} = 3/9[1/3 \;\; 2/3 \;\; 0 \;\; 0 \;\; 0] + 6/9[0 \;\; 0 \;\; 0 \;\; 7/9 \;\; 2/9] = [9/81 \;\; 18/81 \;\; 0 \;\; 42/81 \;\; 12/81] \;\; \clubsuit$$

In the general case the evaluation of the absorbing probabilities can be more complex, an issue postponed to Chaper 2.

1.10 Broad sense stationarity

Limit (1.112) does not always exist, as we have seen in Example 1.131, because, although the chain is irreducible, its asymptotic behavior is periodic.

Example (1.172)
Consider the case where X has two states and keeps changing state at each step. The one-step transition matrix is such that

$$\mathbf{P} = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} \qquad \mathbf{P}^{2n} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} \qquad \mathbf{P}^{2n+1} = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} \tag{1.173}$$

so that a limit of $\mathbf{P}^n$ never exists. We also have

$$\mathbf{\Pi}(2n) = \mathbf{\Pi}(0), \qquad \mathbf{\Pi}(2n+1) = 1 - \mathbf{\Pi}(0). \tag{1.174}$$

The chain $X(n)$ is periodic, and therefore does not have an asymptotic distribution in the ordinary sense, except in the case where $\mathbf{\Pi}(0) = [0.5; 0.5]$. However, even in this case, each realization of the process is periodic and the evolution of the phase depends on the initial conditions. Nevertheless, if we solve the balance equations, we always find $\mathbf{\Pi} = [0.5; 0.5]$.♣.

Let define the *generalized asymptotic distribution* as

$$\mathbf{\Pi}^{**} = \lim_{m\to\infty} \frac{1}{m} \sum_{n=0}^{m-1} \mathbf{\Pi}(n), \tag{1.175}$$

i.e., the arithmetic average on the entire axis, which clearly exists for any periodic distribution. We can immediately prove the following

Property (1.176)
$\mathbf{\Pi}^{**}$ exists for any MC

Property (1.177)
If the ordinary periodic distribution $\mathbf{\Pi}^{*}$ exists, then $\mathbf{\Pi}^{**} = \mathbf{\Pi}^{*}$

Property (1.178)
$\mathbf{\Pi}^{**}$ is a solution of the balance equations.

In fact, from the evolution equation

$$\mathbf{\Pi}(n+1) = \mathbf{\Pi}(n)\mathbf{P},$$

we have

$$\lim_{m\to\infty} \frac{1}{m} \sum_{n=0}^{m-1} \mathbf{\Pi}(n+1) = \left(\lim_{m\to\infty} \frac{1}{m} \sum_{n=0}^{m-1} \mathbf{\Pi}(n) \right) \mathbf{P},$$

which shows that also the asymptotic distribution according to the generalized limit obeys to

$$\mathbf{\Pi} = \mathbf{\Pi}\mathbf{P}.$$

Therefore the unique solution of balance equations, that we find also with periodic chains, must be intended as above.

The generalized asymptotic distribution can be interpreted also in a probabilistic way. Let consider the m time instants such as $0 \le n < m$. If we pick a time instant N at random in this set, we pick time n with probability $1/m$. hence, the distribution we "see" a such time instant can be evaluated by the total probability theorem as

$$\mathbf{\Pi}^{(m)} = \sum_{n=0}^{m-1} \mathbf{\Pi}(n)\frac{1}{m} = \frac{1}{m} \sum_{n=0}^{m-1} \mathbf{\Pi}(n).$$

By (1.175) we see that

$$\mathbf{\Pi}^{**} = \lim_{m\to\infty} \mathbf{\Pi}^{(m)}.$$

Hence, we have the following

Property (1.179)
$\mathbf{\Pi}^{**}$ *can be interpreted as the distribution seen in a time instant picked at random on whole axis.*

The above interpretation face the problem of defining an uniform distribution over the entire time axis, for which we have

$$P(N = n) = \lim_{m\to\infty} \frac{1}{m} = 0,$$
$$P(N < n) = \lim_{m\to\infty} \frac{n}{m} = 0.$$

The probabilities above show that we have zero probability to pick a finite time instant. Which means that the time we select with this procedure is always at infinity. This explain the term asymptotic given to $\mathbf{\Pi}^{**}$. The point taken with the above explained process is called *Random Inspection Point* (RIP).

This interpretation, with respect to Example ex: 3.3, shows that we have 0.5 probability of selecting an even, or odd, time instant. Therefore the generalized asymptotic distribution is $[0.5 \quad 0.5]$, as attained by solving the balance equations.

In the following we assume that the asymptotic distribution always exists, always referring to the generalized asymptotic distributions.

Example (1.180)
The binary chain $X(n)$ alternates sojourn times in states 0 and 1. These times are constant and equal to 3 units in one and 2 units in zero. Find the asymptotic distribution with the RIP method.

The process is periodic with a period of 5 time units. The RIP always lies uniformly within a period and the probability that lies in either states is proportional to the sojourn time in these states. Therefore, we have

$$\pi_0 = 2/5 \qquad\qquad \pi_1 = 3/5 \qquad \clubsuit$$

The generalized distribution has another interesting interpretation.

Theorem: (1.181)
Denoted by $S_k(n)$ the cumulative amount of time spent by the process in state k up to time $n-1$, for asymptotically stationary processes we have

$$\pi_k = \lim_{n\to\infty} \frac{E[S_k(n)]}{n}. \tag{1.182}$$

In fact, if we denote by $Z_k(m)$ the RV that is 1 if the process is in k at time m, (*i.e.*, $X(m) = k$), and zero otherwise, then $E[Z_k(m)] = \pi_k(m)$ and we have, using the broad-sense asymptotic convergence,

$$\pi_k = \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \pi_k(m) = \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} E[Z_k(m)] = \lim_{n\to\infty} \frac{E\left[\sum_{m=0}^{n-1} Z_k(m)\right]}{n} = \lim_{n\to\infty} \frac{E\left[S_k(n)\right]}{n}. \qquad \clubsuit$$

Theorem 1.181 shows that π_k can be interpreted as the percentage of the total average time the chain spends in k.

1.11 Irreducible Markov Chains with infinite state space

We refer to the case of an irreducible finite scalar chain $X(n)$, such as the random walk between barriers 0 and N. In this case, when N becomes unbounded, the distribution at time n can be evaluated as the limiting process:

$$\lim_{N\to\infty} \mathbf{\Pi}(n).$$

However, when we look for the asymptotic distribution we must evaluate

$$\lim_{n\to\infty} \lim_{N\to\infty} \mathbf{\Pi}(n).$$

We can show, but the proof requires advanced tools, that

$$\lim_{n\to\infty} \lim_{N\to\infty} \mathbf{\Pi}(n) = \lim_{N\to\infty} \lim_{n\to\infty} \mathbf{\Pi}(n), \tag{1.183}$$

so that the limit process over N can be taken even for the asymptotic distribution.

However, while in the finite case a proper distribution always exists, when the state space becomes unbounded it can happen that the above limit, which still exists, which is unique and independent from initial conditions for irreducible chains, is such that its components are all zero. This is because, with an infinite number of states, Property (1.133) of irreducible chains do not necessarily hold. It may happen, in fact, that, although any state can be reached by any other, as required by the irreducible property, it is not necessarily reached with probability one. This because the path on the states can take the chain toward infinity and never come back, as it can happen in the unbounded Random Walk. Or if it is reached with probability one it may happen that the average time to reach it is not finite. The irreducible property only implies that a given behavior is common to all states. We can resume this behavior in the following

Property (1.184)
The states of an irreducible and infinite Markov chain are all reached either with probability one or all with probability less than one.

Property (1.185)
When the states of an irreducible and infinite Markov chain are reached with probability one the average time to the visit is either finite or infinite for all states.

The properties above do influence the asymptotic behavior of irreducible chains, since, if a state is not visited with probability one, is asymptotically not visited at all and its probability is zero. The fact that all the elements of the asymptotic distribution are zero means that $X(n)$ grows without limit as n increases, and after an infinite amount of time it can not be found in any finite state j. In other words, the chain has gone to infinity and non-zero probability is assumed only by states at infinite. Therefore:

Property (1.186)
the elements of the unique asymptotic distribution of an infinite, homogeneous and irreducible MC are either all greater than zero or are all zero.

The asymptotic distribution is still obtained by solving the balance equations. However, unique non-zero solution, and the proportional ones, that are found in this way must be normalized by

$\sum x_i = 1$. Here, it may happen that the summation solution $\{x_i\}$ does not converge, i.e., $\sum x_i = \infty$. In this case the only compatible distribution is $\pi_i = 0, \ \forall i$.

In the cases of interest, such as the random walk of Example 1.156, where the space is one-dimensional, the equilibrium solution, if it exists, can be derived from that of finite state space by taking the limit $N \to \infty$.

In this way (1.157) becomes

$$\begin{cases} \pi_i &= (1 - p/q)(p/q)^i, \qquad i \geq 0 \quad p < q \\ \pi_i &= 0, \qquad\qquad\qquad\quad i \geq 0 \quad p \geq q. \end{cases} \tag{1.187}$$

We explicitly note that, in the case $p \geq q$, the only possible equilibrium distribution is the one identically zero, showing that, in that case, a stationary distribution can not exist. This behavior is different from what we have with finite state space, where a stationary distribution, not identically zero, always exists.

We also note that with an infinite number of states the solution becomes more complex, as numerical solutions are excluded and analytic ones can be found only in a few lucky cases.

1.11.1 Non-irreducible chains with infinite state space

With an infinite state space even the transient subset can have an infinite number of states and Property 1.170 non longer holds. In fact, it can happen that the chain stay forever in the transient subset, and the asymptotic probability of reaching any finite state is zero. More on this is given in Chapter 2.

1.12 Generalizations

We remark that many definitions we have provided in the previous sections hold for general processes, such as the broad sense stationarity defined in the previous section, and Theorem 1.181. However, concepts such as homogeneity and irreducibility are more complex to define for non markovian processes. Other formulas still hold, but are not as useful as in markovian processes. An example is given by balance equations. These come from relation (1.108), here rewritten

$$\mathbf{\Pi} = \mathbf{\Pi}\mathbf{P}.$$

It holds for any process that is asymptotically stationary, and we will use it in future sections, but we must note that also matrix $\mathbf{P}$ refer to the stationary second order distribution, or asymptotic second order distribution when they exist:

$$p_{jk} = \frac{\pi_{jk}}{\pi_j}.$$

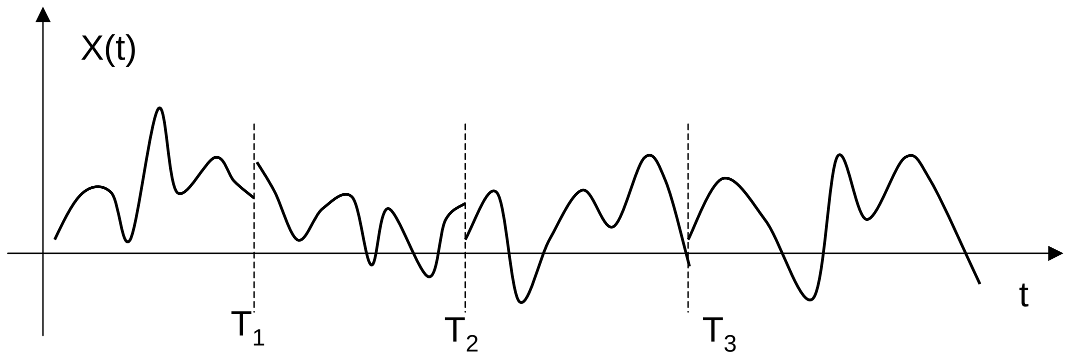


Figure 1.19: *Sample of a regenerative process.*

However, to get such second order distributions, in general we must get first the stationary distribution of any order.

Also the content of Section 1.10, together with the definition of the generalized asymptotic distribution holds for any process. In particular, Theorem 1.181 is valid for any process that is asymptotically stationary.

1.13 Regenerative Processes

A process $X(t, \alpha)$ is called **regenerative** if we can find a sequence of instants $t = t_i(\alpha)$, RVs', at which times the process renews, i.e., it evolves *independent* of the past, and, starting in those instants, presents the *same* probabilistic description, for each value of index i.

Note that two RVs X_1 and X_2, taken at instants t_1 and t_2, are statistically independent only if the two instants belong to different regenerative intervals, otherwise they are not independent.

The above definition implies that if the origin is fixed in one of the regenerative instants t_i, and we considers the process until the next instant of regeneration, for different i we obtain processes $Y_i(t, \alpha)$ independent and equally distributed, that is, they evolve with the same probability law.

In turn, the above implies that the RVs' $C_i = T_{i+1} - T_i$ constitute a sequence of RVs' described by a unique pdf, so we can refer to a single RV C, the interval of regeneration.

The regenerative processes are classified in

- **positive recurrent** if $E[C] < \infty$
- **null-recurrent** if $E[C] = \infty$
- **non-recurrent or transient** if $\displaystyle\int f_C(x)dx = \alpha < 1$, i.e., if RV C is degenerate, i.e., it

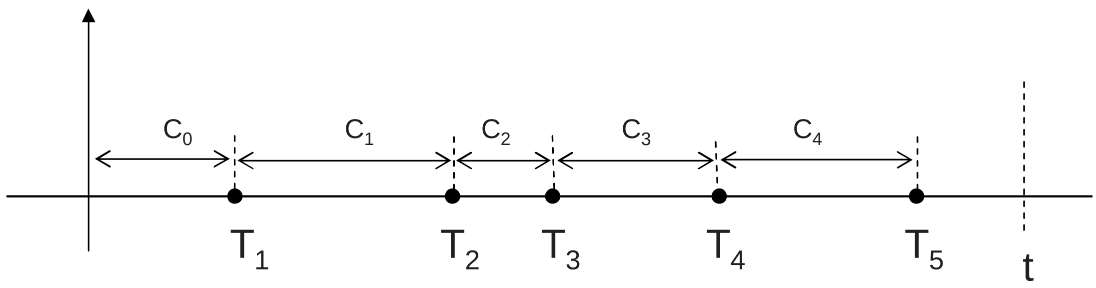


Figure 1.20: *Examples og regeneration points.*

assumes the infinite value with a non zero probability $1 - \int f_C(x)dx$.

Example *Random Walk* (1.188)
With reference to the random walk of Example 1.75 the time instants where the random walk returns to state 0 are regeneration instants. In fact, once you get into that state, the past does not matter and the future evolves always in the same way. We will later show that

- if $p \neq q$ the chain is non-recurrent (the return to zero has probability less than one)
- if $p = q$ the chain is null-recurrent (the return to zero has probability one, but the average return time is infinite)

The following results is basic for positive recurrent processes:

Theorem: (1.189)
In a positive recurrent regenerative process the number $N(t)$ *of regenerative points in* $[0;t]$ *is such that*

$$\lim_{t\to\infty} P(N(t) \geq k) = 1 \tag{1.190}$$

for any finite k. (This is the probabilistic way to say that $N(t)$ increases and goes to infinity as t so does.)

Proof
The following holds true

$$P(N(t) \geq k) = P(T_k < t) = 1 - P(T_k \geq t) = 1 - P(\sum_{i=0}^{k-1} C_i \geq t). \tag{1.191}$$

We now use the *mean value inequality*, that holds true for positive RVs with average a:

$$P(X \geq a) \leq \frac{E[X]}{a}. \tag{1.192}$$

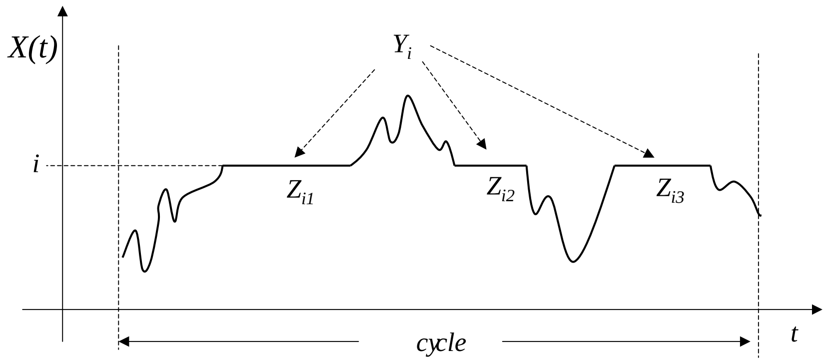


Figure 1.21: Sample of a rigeneration cycle showing the relationship between sojourn times Z_{ij} in state i and the global time Y_i in state i in a rigeneration cycle.

The above, applied to RV $\sum_{i=0}^{k-1} C_i$ yields

$$P(N(t) \geq k) \geq 1 - \frac{kE[C]}{t},$$

which prove the thesis.♣

We recall that, if the process is non-recurring, there is a non-zero probability α that RV C is infinite, which means that with the same probability the process does not regenerate. Therefore, as t grows, sooner or later the process meets this event and does not renew. We have, in fact

Theorem: (1.193)

In a non-recurring regenerative process the number N of regenerative instants is a finite RV, that is,

$$\lim_{k\to\infty} P(N < k) = 1 \tag{1.194}$$

Proof

denoted by $\alpha < 1$ the probability that the cycle is of finite length, N is geometric RV, i.e.,

$$P(N \geq k) = \alpha^k,$$

and

$$\lim_{k\to\infty} P(N < k) = 1 - \lim_{k\to\infty} \alpha^k = 1 \clubsuit$$

The previous theorems describe what happens to the number of regeneration points when t grows in the two cases positive recurrent and non-recurrent. Nothing we can say in the case of null recurrent precesses.

A fundamental results concerning regenerative processes is given by the following

Theorem: (1.195)

A positive recurrent regenerative process always presents a unique asymptotic distribution, independent of initial conditions, given by:

$$\pi_r^* = \frac{E[Y_r]}{E[C]}, \tag{1.196}$$

where Y_r is the RV cumulative sojourn time in state r within a cicle and

$$E[C] = \sum_r E[Y_r] \tag{1.197}$$

is the average cycle time.

The existence, uniqueness and independence from initial conditions come from Theorem 1.189, which states that in the limit we surely have at least a regeneration point. Furthermore, the fact that the cycle time is finite in the average does not allow the process to go to infinity. For proof of (1.196) we refer to Figure 1.21, which shows the relationship between between sojourn times Z_{ij} in state i and the global time Y_i in state i in a rigeneration cycle. The proof is based on result 1.182 and on the assumption that any interval $[0, m]$ can be approximated by the sum of N of cycles, with an error percentage that vanishes as m increases:

$$m \simeq \sum_{i=1}^{N} C_i,$$

where N is the number of cycles in $[0; m]$. This number, because of Theorem 1.189 increases with m. (Note that m is a number, while $\sum_i C_i$ is a RV, but a more precise definition of approximation, in this case, is left to further readings). In the same way we can write

$$E[S_r(m)] \simeq \sum_{i=1}^{N} E[Y_{r,i}] = NE[Y_r],$$

as the average sojourn time in r at each cycle is identical in each cycle. Applying now 1.182 we have

$$\pi_r^* = \lim_{m\to\infty} \frac{E[S_r(m)]}{m} = \lim_{N\to\infty} \frac{NE[Y_r]}{\sum_{i=1}^{N} C_i} = \lim_{N\to\infty} \frac{E[Y_r]}{\sum_{i=1}^{N} C_i/N} = \frac{E[Y_r]}{E[C]}.$$

In the last passage above we have exploited the law of large numbers for independent RV C, which assures

$$\lim_{N\to\infty} \frac{\sum_{i=1}^{N} C_i}{N} = E[C], \tag{1.198}$$

where the convergence is taken with probability one. ♣

Theorem 1.195 shows the surprising result that the asymptotic distribution of the cited process depends *only* on the mean values of times Y_r and not on their distribution, extending the result (1.182). In particular, it generalizes a property that holds for periodic deterministic processes. It

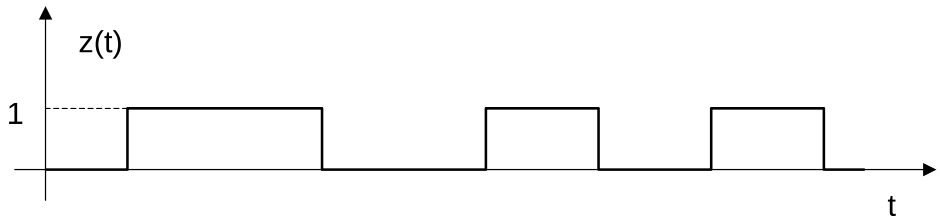


Figure 1.22: *Traffic sample of a single source.*

has great relevance in theory, but, unfortunately, it is of little use in practice since, except in some simple cases, is very difficult to directly get the average times involved.

All the relevant regenerative processes are such that $E[Y_r] > 0$, otherwise state r has no meaning. This leads to the

Corollary (1.199)
In positive recurrent regenerative process all the elements π_r of the asymptotic distribution are positive

Example (1.200)
The continuous-time regenerative process $Y(t)$ enters cyclically into the states 0, 1, 2, where it remains a fraction of time proportional to 1, 2 and 3 respectively. Find the asymptotic distribution.

We apply Theorem 1.195, which shows that

$$\pi_0 = K \times 1 \qquad \pi_1 = K \times 2 \qquad \pi_2 = K \times 3$$

where K is a constant that we find using the congruence equation $\sum_i \pi_i = 1$. We get $K = 1/6$ and

$$\pi_0 = 1/6 \qquad \pi_1 = 1/3 \qquad \pi_2 = 1/2 \quad \clubsuit$$

Example (1.201)
The traffic generated by a source is represented by a binary process $Z(t), (Z = 0, 1)$ that represents the activity, $Z = 1$ representing the active state, *and $Z = 0$ representing the* idle state *(Figure 1.22). A regenerative source model is the one that assumes regenerative instants when the source becomes idle.*

For positive-recurrent regenerative traffic, where the average duration of the idle period is m_Y and the average duration of the activity period is m_X, the average traffic in stationary (asymptotic) conditions is, by Theorem 1.195,

$$S = E[Z] = \pi_1 = \frac{m_X}{m_X + m_Y} \tag{1.202}$$

Corollary (1.203)
In a positive recurrent regenerative process $Z(t)$ the average of $E[Z(t)]$, in stationary conditions, can be derived as

$$E[Z(t)] = \sum_r r\pi_r = \sum_r r\frac{E[Y_r]}{E[C]} = \frac{E\left[\int_C Z(t)dt\right]}{E[C]}, \tag{1.204}$$

where the integral is taken on a cicle.

In fact, we deterministically have

$$\sum_r rY_r = \int_C Z(t)dt.$$

With respect to Theorem 1.195, we can say nothing general if the process is not positive recurrent. For example, it may happen that the asymptotic distribution is zero for some states or for all states (we see below that the latter happen with non-positive recurrent MC).

1.13.1 Regenerative processes and Markov Chains

There is a strong relationship between Markov Chains and Regenerative processes, which is stated in the following

Property (1.205)
Homogeneous Markov Chains are regenerative, and the regeneration instants are those where the chain enters a state r, arbitrarily preselected.

In fact, whenever you enter a state r, the evolution of the chain is independent of the past (markovian property) and with the same probabilistic law (homogeneity).

The above property shows that a single chain may be seen as different regeneration processes, according to the choice of state r.

In discrete-time MCs regeneration points are adjacent as long as the chain does not leave state r. In any cases we always have

$$E[Y_r] = 1, \qquad\qquad E[C] = m_{rr}, \tag{1.206}$$

where m_{rr} is called the mean return time to r.

In accordance with the classification of regenerative processes, a state r of a CM is called recurrent if the probability of return to r is 1, and non-recurrent otherwise. It is positive recurrent if m_{rr} is finite, otherwise it is null-recurrent. We have

Theorem: (1.207)
The asymptotic distribution of a homogeneous MC is such that if r it is positive recurrent we have

$$\pi_r^* = \frac{1}{m_{rr}} > 0, \qquad \forall r, \tag{1.208}$$

otherwise, if it is null recurrent or non-recurrent we have

$$\pi_r^* = 0, \qquad \forall r. \tag{1.209}$$

Result (1.208) comes from Theorem 1.195 using (1.206). If r is non-recurrent, asymptotically it will never be reached , the fraction of time it spends in state r is zero and, therefore, by Theorem

1.181, we get $\pi_r^* = 0$. The same holds if r is null-recurrent because, again, the fraction of time it spends in state r is zero.

Example (1.210)
Let consider a Random walk with a reflecting barrier in zero and an absobing barrier in N. Then State N is positive recurrent with $m_{rr} = 1$, whereas the others are all non-recurrent. Hence, asymptotically we have

$$\pi_N = 1, \qquad \pi_i = 0, \; i \neq N. \qquad \clubsuit$$

Focusing on irreducible Markov Chains, the very same argument that showed Properties 1.133, 1.184 and 1.185 lead us to the following

Property (1.211)
the states of a homogeneous and irreducible Markov Chain are all of the same type, i.e., positive recurrent, null recurrent or non-recurrent, and the entire chain is named with the same type.

In particular, by Property 1.133, we have

Property (1.212)
A finite homogeneous and irreducible Markov Chain is positive recurrent.

Finally, Theorem 1.207 can be used to state the following

Theorem: (1.213)
A infinite homogeneous and irreducible Markov Chain is positive recurrent if and only if a non-zero equilibrium distribution exists, otherwise it is not positive recurrent.

Note that this theorem provides us with a means to determine the nature of an MC.

Example (1.214)
Referring to the Random Walk studied in Example 1.167, we see that for $p < q$ the equilibrium distribution exists and, therefore the chain is positive recurrent, meaning that the average return time to any state is finite. Otherwise, the equilibrium distribution does not exists and, therefore the chain is not positive recurrent, meaning that the average return time to any state is infinite.

In continuous-time MCs we can not adopt definitions (1.206) because the single time step reduces to zero. Therefore, our entrance into state r is considered only when the chain comes from other states, so that we have

$$E[Y_r] = \frac{1}{-q_{rr}}, \qquad E[C] = m'_{rr}, \tag{1.215}$$

where m'_{rr} is the new definition of the average return time.

1.14 The transition chain

Let refer to continuous-time chains such as chain $X(t)$ in Figure 1.23 (the same holds true for discrete time but a change of notation). If we only look at transition instants we get the following

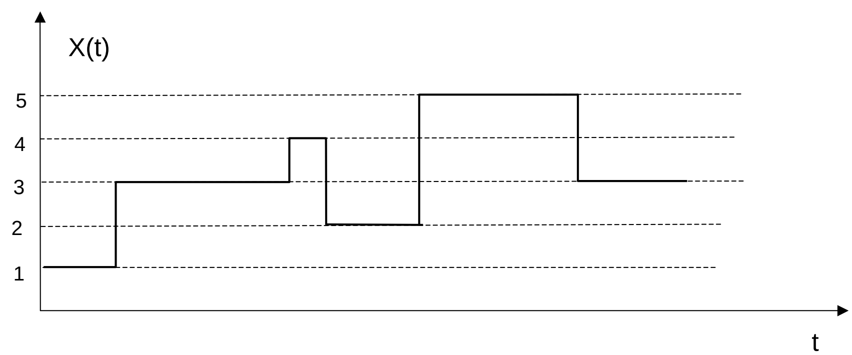


Figure 1.23: *Sample of continuous-time chain $X(t)$.*

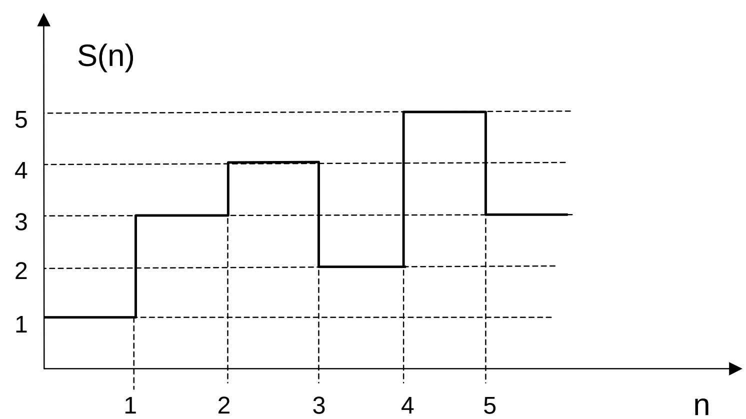


Figure 1.24: *Transition chain $S(n)$ derived from continuous-time chain $X(t)$ of figure 1.23.*

transition probabilities *conditional* to the occurrence of a state transition

$$p'_{jk} = \frac{q_{jk}}{\sum_{r \neq j} q_{jr}} = \frac{q_{jk}}{-q_{jj}}, \qquad j \neq k. \tag{1.216}$$

The transition probabilities above can be seen as those of the discrete-time *transition chain* $S(n)$ in Figure 1.23, derived from $X(t)$ with the same transitions but with sojourn times equal to one time unit. Relation (1.216) can be reversed in the following way:

$$q_{jk} = p'_{jk}(-q_{jj}) = \frac{p'_{jk}}{E[Z_j]}, \qquad j \neq k, \tag{1.217}$$

being Z_j the sojourn time in state j.

Example (1.218)
Given a continuous-time random walk in $[0;\infty]$ with parameters λ and μ, its transition chain has parameters

$$p_{01} = 1, \qquad p_{i,i+1} = \frac{\lambda}{\lambda+\mu},\ i > 0, \qquad p_{i,i-1} = \frac{\mu}{\lambda+\mu},\ i \geq 0.\ \clubsuit$$

We can also directly relate the distribution of the two chains. If we denote by π_i and ν_i the distributions of $X(t)$ and $S(n)$ chains respectively, we have

Theorem: (1.219)

$$\pi_r = \frac{\nu_r E[Z_r]}{\sum_i \nu_i E[Z_i]}. \tag{1.220}$$

In fact, starting by theorem (1.196) for $X(n)$ we have:

$$\pi_r = \frac{E[Y_r]}{\sum_i E[Y_i]}, \tag{1.221}$$

and for $S(n)$ we have

$$\nu_r = \frac{e_r}{\sum_i e_i}, \tag{1.222}$$

where e_r is the average number of times the chain enters state r, which represents time Y_r in the transition chain (each time it enters a state it stay there only a unit of time). Observing that

$$E[Y_r] = e_r E[Z_r],$$

from (1.221) we then have

$$\pi_r = \frac{e_r E[Z_r]}{\sum_i e_i E[Z_i]}. \tag{1.223}$$

The thesis is proven by replacing in the above e_r with ν_r derived by (1.222).$\clubsuit$

1.15 Semi-Markov chains

Semi-Markov chains are chains whose transition chain is markovian, while the sojourn time has a pdf that is not memoryless (i.e. geometric or negative exponential). These chains are not markovian, because the sojourn time has memory; nevertheless, they are regenerative chains. The limiting theorem for regenerative processes (1.196) shows that the distribution depends only on the averages of the sojourn times in states, but not on their pdf. Therefore, in order to evaluate the distribution of the Semi-Markov chain we can replace the pdf of the sojourn time with a negative exponential of the same average. Then we have proved the following

Theorem: (1.224)
The distribution of a Semi-Markov chain can be obtained exactly as if it were a Markov Chain, where the transition rates are derived by (1.217).

1.16 Ergodic processes

Theorem 1.182 shows that the element of the asymptotic distribution π_k is the percentage of the average total time the chain spends in state k. However for some processes we also have

$$\pi_r = \lim_{n\to\infty} \frac{E[S_r(n)]}{n} = \lim_{n\to\infty} \frac{S_r(n)}{n}, \tag{1.225}$$

where the second limit, which involves RV $S_r(n)$, must be taken in the sense specified below. The 1.225 says that in order to measure π_k we do not need to take the average over many samples, but it suffices the $S_r(n)$, measured over a single realization. Processes with such a property are called *ergodic*.

In general the measure of a probability $P(A)$ (and therefore of a distribution) can be made, thanks to the law of large numbers, by N independent experiments and counting the number of times N_A where A occurs. In order to measure the probability $\pi_r(n)$, we must perform N experiments that provide N process samples, and then evaluate the frequency of state r at time n, over a sufficiently large N, i.e.:

$$\pi_r(n) = \lim_{N\to\infty} \frac{N_r(n)}{N}. \tag{1.226}$$

It should be noted that the limit (1.226) can not be taken in the ordinary sense since it operates on a RV. The theory defines several criteria of convergence for random variables, here omitted; the strongest of these criteria is the convergence with probability 1.

If the process is stationary then (1.226) does not depend on n and therefore the measurement can be carried out with an arbitrary n. If we look for the asymptotic distribution we have to look for a far away n and then proceed as above.

Measure (1.226) needs an infinite number of samples of the process even if stationary. With ergodic processes a measure of this kind can be carried out on a single realization.

Definition: A process $X(t, \alpha)$ is called *ergodic* if, being stationary or asymptotically stationary, all its probabilistic measures (averages, distributions of any order, etc.) are attainable as time averages on a single sample.

If the process has a discrete state space r, then its distribution can be obtained as

$$\pi_r = \lim_{t\to\infty} \frac{S_r(t)}{t}. \tag{1.227}$$

Extensions to other probability measures can be given in a similar way.

Law (1.227) represents a generalization of the law of large numbers, which holds for independent RV, to RV that are *not* independent (memory). For non-ergodic processes limit (1.227) often exists, but yields a random variable, that changes from realization to realization. The ergodicity is, therefore, a very important property as it allows the measures of probabilities through a single experiment, for example with the simulations. By (1.227) π_r is interpreted as the percentage of time that the process spends in state r.

Unfortunately, it is not easy to determine when a processes is ergodic. For example, we immediately see that a *purely random, and stationary process* is ergodic, because to it applies the law of large numbers. Qualitatively, we can say that this property is related to the number of regenerations (loss of memory) the process encounter over time, so that the RVs taken in two instants sufficiently apart become independent. In this way pieces of a single realization sufficiently far apart become independent just as if they belonged to different samples, and the law of large numbers applies.

We may also say that with ergodic processes the probabilistic characteristics of the whole process are enclosed in a single sample. In fact, you could say that, with the due caution, all samples can be derived by delaying for time amount, possibly infinite, one single sample.

By what we have said, it appears that processes whose asymptotic distribution does depend on initial conditions can not be ergodic. On the other side we have

Theorem: (1.228)
regenerative positive-recurrent processes are ergodic.

This comes from Theorem 1.189, that assures that with such processes the number of regenerations increases with time, so that after an infinite amount of time we have an infinite amount of regenerations, that is, an infinite amount of samples.

More formally, we observe that, like in Theorem 1.195, any interval $[0;t]$ can be approximated by the cumulated amount of cicles durations:

$$t \simeq \sum_{i=1}^{N} C_i,$$

where N is the number of cicles (actually a RV) min $[0;t]$. In the same way we can approximate

$$S_r(t) \simeq \sum_{i=1}^{N} Y_{r,i},$$

being $Y_{r,i}$ the total amount of time spent by the process in state r during i-th cycle. We then have

$$\lim_{t\to\infty} \frac{S_r(t)}{t} = \lim_{t\to\infty} \frac{\sum_{i=1}^{N} Y_{r,i}}{\sum_{i=1}^{N} C_i} = \lim_{N\to\infty} \frac{\sum_{i=1}^{N} Y_{r,i}/N}{\sum_{i=1}^{N} C_i/N} = \frac{E[Y_r]}{E[C]}.$$

The result comes from the exchange of t and N (Theorem 1.189), and the law of large numbers (1.198) applied to RVs $Y_{r,i}$, that are independent because belonging to different cycles. Similarly for RVs C_i. For Theorem 1.195 the above result equals π_r.

We again recall that the limits above must be taken with respect of convergence criteria for RVs.♣

From this theorem we also have, for obvious reasons,:

Corollary (1.229)
Homogeneous, irreducible and positive recurrent MCs are ergodic.

The ergodic property, like other properties we have demonstrated for positive-recurrent regenerative processes, applies to a more extended class of processes in which the memory reduces in times, even though not abruptly as it happens with regenerative processes. Unfortunately, for such processes, with some exceptions, there are not general rules for detreminingn the ergodicity.

1.16.1 Measures and simulations

Ergodicity simplifies the way of getting asymptotic measures when they exist. However, since time span n on which to take averages is finite, the problem exists to determine how long n must be for a safe measure. Statistics gives us an answer when dealing with the law of large numbers for independent RV, or with different independent samples of a process.

Referring to positive recurrent processes, the roles of independent variables is taken by independent cycles. In other word, we must observe a sufficiently great number of independent cycles in order to exploit (1.225). Therefore, during the measure, we must look for the occurrence of regeneration points.

As an example, let assume we refer to Random Walk of Example 1.167. We have a sample and we want to determine the nature of this process, and its asymptotics if they exists. A trivial solution would be to measure probabilities p and q, and decide applying theoretical results. That is we must first decide among $p \gtreqless q$. here we see the first problem occur when p and qv are very close, since this require very long measures to get the needed precision.

However, here the Random Walk is used to exemplify more complex processes, whose behavior depends from a greater number of variables, as it happens for generalized Random Walk, or processes with longer memory. Here the procedure is to look for regeneration points and make sure they grows with time, according to Theorem 1.189, otherwise the process is non recurrent. Also, we must track the arithmetic average of regeneration intervals, to determine if the process is positive recurrent, and finally we can apply (1.225). Referring to our Random Walk example, and taking as regeneration points the instants where the chain enters state zero, we have (Theorem 1.207):

$$E[C] = m_{00} = \frac{1}{\pi_0} = \frac{1}{1-p/q}.$$

From the above we see that the regeneration interval m_{00} tends to infinity as $p \to q$, meaning that the needed measure may become extremely long in these circumstances.

1.17 Problems for solution

P.1.1 Find the probability that k Poisson Arrivals at rate λ lie within an interval whose width is negative exponential RV of rate μ.

P.1.2 Two Poisson Arrival streams, of rate λ and μ respectively, merge on the same axis. Determine

a) The rate of the composite flow and probability that an arrival in the composite flow belongs to the first stream;

b) the probability that between two consecutive points of the second flow there are k of the first. (remember that $\int_0^\infty x^k e^{-x} dx = k!$)

P.1.3 Two Poisson Arrival streams, A and B, of rate λ_A and λ_B respectively, merge on the same axis. Determine

a) The distribution of the number of arrivals in $[0;T]$;

b) the probability that an arrival in the composite flow belongs to the first stream;

c) the probability that an A arrival is followed by another A arrival;

d) If case c) occurs, find the pdf of the distance of such arrivals (a conditional pdf)

P.1.4 At a bus stop busses arrive according to a Poisson process of rate μ. Passengers arrive at the bus stop according to a Poisson process with rate λ. A bus arrives at the stop and find zero passengers waiting.

a) find the pdf of the time elapsed since the last bus arrived.

b) and if the number of waiting is one?

P.1.5 Consider Poisson arrival of rate λ. Someone tells us that within interval $[0;T]$ lie n arrivals. Find the the distribution of the number of arrivals M that lies in interval $[0;t], t \leq T$.

P.1.6 Let us take the origin of time axis exactly at the time of a Poisson arrival, where the parameter is λ. Now take time τ. Find the pdf of Z, the distance from τ of the time instant where the last arrival before τ occurred.

P.1.7 Show that from the Markov property we also have

$$P(X_{n+2}, X_{n+1}/X_0, \ldots X_n) = P(X_{n+2}, X_{n+1}/X_n)$$

P.1.8 The binary MC X_n ($x_n = 0, 1$) presents the following one step transition matrices **??**

$$\mathbf{P}(n, n+1) = \begin{vmatrix} 0.3 & 0.7 \\ 0.2 & 0.8 \end{vmatrix}$$

If the chain starts in $X = 0$ at time $n = 0$, find the first order distribution at times $n = 1$ and $n = 2$. Find also the joint second order distribution at times $1, 2$ and the third order distribution at times $1, 2, 3$.

P.1.9 A Markov Chain presents the following one-step transition matrix at all times n.

$$\begin{vmatrix} 0.4 & 0.6 & 0 \\ 0 & 0.5 & 0.5 \\ 0.3 & 0.3 & 0.4 \end{vmatrix}$$

If the chain starts in $X = 0$ at time $n = 0$, find

a) the first order distribution $\mathbf{\Pi}(n)$, the transition matrix $\mathbf{P}(0, n)$, and the second order distribution $\mathbf{\Pi}(n, n+1)$ (all terms) all at time $n = 2$;

b) the probability $B(n)$ that the chain is either in state 1 or in state 2 at time $n = 2$; the first order distribution $\mathbf{\Pi}(n)$ at time $n = 4$ knowing that the chain is in state 2 at time $n = 2$; the same, knowing that the chain is in the space subset $(1, 2)$ at time $n = 2$;

P.1.10 Two urns contain each N balls, N of which are blue and the others red, subdivided at random between the urns. At each instant, draw a ball from both urns and, with probability $1/2$ exchange urn. The number of red balls in the first urn, $X(n)$ is Markov chain. Assuming $N = 2$, we write the one-step transition matrix and find the first order distribution at times $n = 1, 2, 3$ assuming that at the time $n = 0$ all the red balls are in the first urn.

P.1.11 (The ruin problem) Two gamblers bet against each other, having probability $1/2$ to win. Each time they bet one unit and the game ends when one of the two has not capital left. The capital of the first player is a Markov chain. In case the players start with capitals 1 and 2 respectively, write the one-step transition matrix and find the first order distribution at times $n = 1, 2, 3$.

P.1.12 Process $Y(n)$ is defined by:

$$Y_{n+1} = Y_n \oplus Y_{n-1} \oplus X_{n+1}$$

where $\oplus$ is the binary summation and X_i is a sequence of binary $[0; 1]$ and independent RVs' with $P(X_i = 1) = p$. Show that $Y(n)$ is not (first order) Markov. The chain becomes Markov by re-defining states. Find the new state variable and draw the corresponding state diagram.

P.1.13 Process X_n is a binary MC $[0;1]$. Say whether process $Y(n)$:

$$Y_n = X_n \oplus X_{n-1}$$

is a MC.

P.1.14 In a digital communication system symbols are continuously transmitted and p is the probability that there is a receiving error. The control system declares the system out of service ($X(t) = 0$) if N consecutive errors are discovered. On the other side, if the system is down, it is declared operational ($X(t) = 1$) after M correct and consecutive symbols are received. Show that the binary state $X(t)$ is not Markov, and define new states such that the new process $Y(t)$, derived by $X(t)$, is Markov. Finally, draw the state transition diagram for $Y(t)$.

P.1.15 Given the following MCs, $X_{n+1} = X_n \oplus V_n$ and $Y_{n+1} = Y_n \oplus W_n$ Where $\oplus$ means binary summation, and the following binary and independent RVs V_n e W_n, with distribution respectively equal to p, $1-p$, and q, $1-q$. Check whether process $Z_n = X_n \oplus Y_n$, is markovian.

P.1.16 Let $X(n)$ be a random walk with a reflecting barrier in zero, with parameters p and q, $p+q < 1$, where the residence period in each state can not exceed two time units. At the first time unit in state i transitions probabilities are p and q, as in the original walk. At the second time unit the only transition is toward state $i+1$ with probability one. Chain $X(n)$ is not Markov, but extending appropriately the state space, it is possible to define a Markov chain $Y(n)$ able to represent $X(n)$. Draw the state transition diagram for $Y(t)$.

P.1.17 Let $X(n)$ be a random walk with a reflecting barrier in zero, with parameters p and q, $p+q < 1$, where we operate the following modifications. When the chain enters state zero, it remains there for two further time units (i.e., three time units in total), after wich it goes to state 1 with certainty. Chain $X(n)$ is not Markov, but extending appropriately the state space, it is possible to define a Markov chain $Y(n)$ able to represent $X(n)$. Draw the state transition diagram for $Y(n)$.

P.1.18 Check whether distribution $[2/9; 7/9]$ is a stationary distribution for the chain in Problem P.1.8

P.1.19 Check whether distribution $[1/6; 4/6; 1/6]$ is a stationary distribution for the chain in Problem P.1.9

P.1.20 Let $\mathbf{P}$ be the one-step transition matrix of the homogeneous MC $X(n)$. Take the chain $Y(n) = X(-n)$, shows it is a MC and write its one-step transition matrix $\mathbf{P}'$ (i.e., derive the new p'_{jk} as function of the old p_{jk} and ...).

P.1.21 The binary MC, X_n, ($x_n = 0, 1$), presents the following one step transition matrix

$$\mathbf{P}(n, n+1) = \begin{vmatrix} 1/4 & 3/4 \\ 1/2 & 1/2 \end{vmatrix}, \qquad \forall n.$$

If the chain starts in $X = 0$ at time $n = 0$, find

a) the first order distribution $\pi_j(n)$ (all terms) at times $n = 2$ and $n = \infty$;

b) the second order distribution $\pi_{jk}(n-1, n)$ (all terms) at times $n = 2$ and $n = \infty$;

c) the vector of the initial conditions $[\pi_0(0), \pi_1(0)]$, that makes $\pi_j(n)$ stationary at all $n > 0$.

d) the transition matrices $\mathbf{P}(0,2)$ and $\mathbf{P}(0,n)$ with $n \to \infty$ (suggestion: note that $p_{jk}(0,\infty) = ...$).

P.1.22 From a stack of N books customers take one at random, browse it and put it back at the top of the stack. The position on the stack $X(n)$ of a particular book is a MC.

a) Trace the state diagram of the MC

b) find the asymptotic probability that the book is found at the top of the stack.

P.1.23 Find the asymptotic distribution of a random walk with parameters p and q, having reflecting barriers in states $-N$ and N.

P.1.24 Consider a random walk with parameters p and q, having reflecting barriers in states 0 and 2:

a) find all the nine elements of the joint asymptotic distribution $\pi_{jk}(n, n+1)$

b) find all the nine elements of the joint asymptotic distribution $\pi_{jk}(n, n+2)$

c) find all the nine elements of the joint asymptotic distribution $\lim_{r\to\infty} \pi_{jk}(n, n+r)$.

P.1.25 Find the asymptotic distribution of $Y(n)$ in problem P.1.12.

P.1.26 Find the asymptotic probabilities that the system in Problem P.1.14 is out of service (outage).

P.1.27 Find the asymptotic distribution of the chain in problem P.1.16.

P.1.28 Find the asymptotic distribution of the chain in problem P.1.17.

P.1.29 A discrete-time MC has the following transition matrix over states $0, 1, 2, 3, 4$:

$$\mathbf{P} = \begin{vmatrix} 0.2 & 0.8 & 0 & 0 & 0 \\ 0.4 & 0.6 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.8 & 0.2 \\ 0 & 0 & 0 & 0.3 & 0.7 \end{vmatrix}$$

Find

a) the distribution vector at time $n = 2$ in the two cases when it starts at time $n = 0$ in states $X = 0$ and $X = 4$ respectively;

b) the asymptotic distributions in the two cases in a)

c) verify whether the vector $[1/6; 2/6; 0; 3/10; 2/10]$ represents an asymptotic (stationary) distribution of the chain;

P.1.30 A binary MC X_n has the following transition probabilities

$$p_{01}(n, n+1) = p_1, \qquad p_{10}(n, n+1) = q_1, \qquad n = 2k+1$$

i.e., when n is odd, and

$$p_{01}(n, n+1) = p_2, \qquad p_{10}(n, n+1) = q_2, \qquad n = 2k$$

i.e., when n is even. Find the asymptotic distributions in even instants, in odd instants, and in any instant.

P.1.31 Let $X(n)$ be a random walk between states 0 and N where at each state $i > 0$ the process stops for the next step and and the second step goes up or down with probability respectively equal to p and $1-p$. For $i = 0$ in the process stops for a further step and then switch to $i = 1$ with certainty.

P.1.32 A non-homogeneous MC is such that the matrix of transitions in a step $\mathbf{P}(n, n+1)$ is periodic with period equal to 3 starting from steps for $n = 0, 1, 2$, and taking the following values

$$\begin{vmatrix} 0.4 & 0.6 & 0 \\ 0.4 & 0 & 0.6 \\ 0 & 0.4 & 0.6 \end{vmatrix} \qquad \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} \qquad \begin{vmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix}$$

Staring at $n = 0$ in state 0 find

a) the distribution at times 1, 2 e 3.
b) the asymptotic distribution at times $3n$, $3n+1$ e $3n+2$, $n = 0, 1, 2, \ldots$;
c) the asymptotic distribution (i.e., in a RIP).

P.1.33 (3.42) A bidimensional random walk in the discrete time moves on step in one of the four directions North, East, South, West, with probability respectively equal to p, p, q, q. Axis are reflective barriers so that the walk only occurs in the first quadrant (X and Y non-negative integers). Denoted by (i, j) the state, i.e., the position in the walk:

a) Check whether a distribution of the form $\pi_{i,j} = K(p/q)^{i+j}$ satisfies the balance equations, and if so find K.
b) Find the distribution of the walk on the horizontal line, i.e., The marginal π_i;
c) Same as a) and b) when the walk is limited by the reflecting barrier $X + Y = 2$.

P.1.34 (4.1) The rate matrix $\mathbf{Q}$ of a continuous-time MC presents the following non-null elements outside the the main diagonal:

$$q_{i,i+1} = \lambda \qquad\qquad i = 0, 1, 2, 3, \ldots$$

$$q_{i,i-2} = \mu \qquad\qquad i = 2, 4, 6, \ldots$$

Find the equilibrium distribution $\pi_i\ \forall\ i$. Is this chain reversible? Draw some sample paths.

P.1.35 (4.7) A MC is composed of two states, 0 and 1, where 1 is an absorbing state. Find the transient distribution $\pi_0(t)$ e $\pi_1(t)$ assuming that the chain starts in state 0.

P.1.36 (4.9) A process $X(t)$ leaves state 0 with rate μ and proceeds toward the states $1, 2, 3, 4$ successively, and stay in each of them for a period of time of length $i, i \in 1, 2, 3, 4$. After leaving state 4 it returns to state 0, and the cycle re-starts again. Derive, explaining, the asymptotic distribution.

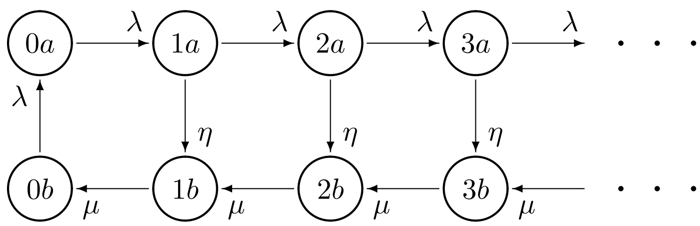
P.1.37 (4.16) A continuous-time binary chain $X(t)$ remains in state 1 a period of time τ and 2τ alternatively, while it still remains a period of time 2τ in state 0.

a) Find the asymptotic distribution;

b) find the average (asymptotic) rates q_{10} e q_{01}.

P.1.38 Consider a time continuous random walk on the positive axis and reflecting barrier at the origin, with rates λ and μ. Assume now that when it enters state 0 it remains there for a constant time T and then goes to state 1. Find the asymptotic distribution.

P.1.39 (4.3) - A continuous-time MC presents the following state diagram:



Find the asymptotic distribution π_{ix}, and the conditions on the parameters such that the asymptotic distribution exists. Check whether the distribution remains the same (up to a constant factor) when the chain is truncated to state N (ie, flux out of Na upward, and flux into Nb downward are canceled.

P.1.40 A random walk process $N(n)$ is such that it moves upward and downward according to a pre-set direction. At each step the direction of the step is changed, before leaving the state, in this way: if the direction was upward it is changed with probability p; if the direction was downward it is changed with probability; q. Upon entering the origin, the direction is immediately reversed. Process $N(n)$ is not markovian, but we can derive its distribution defining a new process $Y(n)$ that is markovian.

a) draw the state diagram of process $Y(n)$;

b) find the asymptotic distribution of $N(n)$ together with the conditions on the parameters such that the asymptotic distribution exists.

P.1.41 A Markov Chain presents the following one-step transition matrix at all times n.

$$\begin{vmatrix} 0.4 & 0.6 & 0 \\ 0.4 & 0 & 0.6 \\ 0 & 0.4 & 0.6 \end{vmatrix}$$

It starts at time 0 in state 0 and at all time instants $n = 3k$, $k = 1, 2, 3, \ldots$, is forced to assume value 0. Find

a) the distribution at times 1, 2 e 3.

b) the asymptotic distribution ($n \to \infty$) at times $3n$, $3n + 1$ e $3n + 2$, $n = 0, 1, 2, \ldots$;

c) the asymptotic distribution (at time n, $n \to \infty$) (i.e., in a RIP).

P.1.42 Three cards are distributed in positions 0, 1, 2. At each step, the pair of positions 0 and 1 is selected with probability 1/3 and the cards exchange positions. Otherwise, the pair 1 and 2 is selected and the cards exchanged. If an Ace is placed in position 0, what is the distribution of its position after two steps? and after an infinite number of steps? What if, at each step, all pairs are selected with the same probability and the cards are exchanged?

P.1.43 A binary MC X_n has the following transition probabilities

$$p_{01}(n,n+1) = 1/2, \qquad p_{10}(n,n+1) = 1/2, \qquad n = 3k$$

$$p_{01}(n,n+1) = 1/4, \qquad p_{10}(n,n+1) = 3/4, \qquad n = 3k+1$$

$$p_{01}(n,n+1) = 1, \qquad p_{10}(n,n+1) = 1, \qquad n = 3k+2$$

Find

a) the distribution at times 1, 2 and 3, starting in state 0 at time 0.

b) the asymptotic distribution ($n \to \infty$) at times $3n$, $3n+1$ e $3n+2$, $n = 0, 1, 2, \ldots$;

c) the asymptotic distribution (at time n, $n \to \infty$) (i.e., in a RIP).

P.1.44 A random walk between zero and infinite has parameters p and q, $p < q$. Find

a) the asymptotic joint probability $\pi_{j,k}(n, n+1)$ (any j, k);

b) the asymptotic joint probability $\pi_{j,k}(n, n+\infty)$ (any j, k);

c) the asymptotic probability of state j when arriving from state $j-1$, and the one when arriving from state $j+1$.

P.1.45 Two discrete-time MC's have the following transition matrices

$$\mathbf{P}_1 = \begin{vmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{vmatrix} \qquad \mathbf{P}_2 = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}$$

A new chain starts at time $n = 0$ in state 0 using the first matrix. Then at time $n = 2$ uses the second matrix, and switches back to the first matrix at time $n = 3$, the second at time $n = 5$, again the first at time $n = 6$ and so on, periodically; Find

a) the distribution vector at times $n = 2, 3, 4$, the transition matrices $\mathbf{P}(0,2)$ and $\mathbf{P}(0,3)$;

b) $\lim_{n\to\infty} P_1^n$;

c) the asymptotic distribution at times $n = 3k$, $k = 0, 1, 2, \ldots$ (note that it can be derived as a limit from $\mathbf{P}(0,3)$.

c) the asymptotic distribution;

P.1.46 A Markov Chain presents the following one-step transition matrix at all times n.

$$\mathbf{P} = \begin{vmatrix} 0 & 0.6 & 0.4 \\ 0.4 & 0.6 & 0 \\ 0.6 & 0 & 0.4 \end{vmatrix}$$

If the chain starts in $X = 0$ at time $n = 0$, find

a) the first order distribution $\mathbf{\Pi}(n)$, and the transition matrix $\mathbf{P}(0, n)$ at time $n = 2$;

b) the first order distribution $\mathbf{\Pi}(n)$, and the transition matrix $\mathbf{P}(0, n)$ at time $n = \infty$;

c) the second order distribution $\mathbf{\Pi}(n-1, n)$ (all terms) at times $n = 2$ and $n = \infty$;

d) the vector of the initial distribution $\mathbf{\Pi}(0)$ that makes the chain stationary at all $n \geq 0$.

P.1.47 Given a two-state $(0, 1)$ discrete-time Markov Chain $X(n)$, $n = 0, 1, \ldots$ with transition probabilities p and q, derive chain $Y(n)$, $n = 1, 2, \ldots$ such that $Y(n+1) = X(n) \oplus X(n+1)$, with $\oplus$ meaning binary summation. Find

a) the first order distribution $\mathbf{\Pi}(n)$ of $Y(n)$ at time $n = 2$ starting at time $n = 0$ in with $X(0) = 0$ and $Y(0) = 0$.

b) the asymptotic distribution $\mathbf{\Pi}$ of $Y(n)$.

c) the asymptotic second order distribution $\mathbf{\Pi}(n, n+1)$, i.e., $n = \infty$.

P.1.48 The hour, from 1 to 12, displayed by a clock changes each time by one hour clockwise with probability 1/2, otherwise it does not change.

a) find the asymptotic distribution, if any, of the hour displayed by the clock.

Consider now the case where the clock, in addition to the step clockwise with probability 1/2, is also allowed to step by one hour counterclockwise with probability 1/3. If it starts at 12 at step $n = 0$,

b) find the distribution of the displayed hour at steps $n = 2$ and $n = 3$;

c) verify whether or not the asymptotic distribution, if any, of the hour displayed by the clock is uniform.

P.1.49 A Markov Chain presents the following one-step transition matrix at all times n.

$$\begin{vmatrix} 0.4 & 0.6 & 0 \\ 0.4 & 0 & 0.6 \\ 0 & 0.4 & 0.6 \end{vmatrix}$$

a) Find the asymptotic (stationary) distribution.

b) Knowing that at time n_0 the distribution of the chain is the stationary one, find the distribution at time $n_0 + 2$.

c) Knowing that at time n_0 the chain is in state 0, find the distribution at time $n_0 + 2$.

P.1.50 A binary non-homogeneous Markov Chain $X(n)$ presents the following one-step transition matrices n.

$$\mathbf{P}(2n, 2n+1) = \begin{vmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{vmatrix} \qquad \mathbf{P}(2n+1, 2n+2) = \begin{vmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{vmatrix}$$

for any $n = 0, 1, 2, \ldots$. Starting at time zero with distribution $[1/2,\ 1/2]$ find:

a) the distribution vector at times $n = 2, 3, 4$, and the transition matrices $\mathbf{P}(0, 2)$ and $\mathbf{P}(0, 3)$;

b) the asymptotic distributions at even time instants, at odd time instants and at any time.

c) the asymptotic joint distribution (all four elements) at two adjacent time instants.

P.1.51 A Markov Chain presents the following one-step transition matrix at all times n.

$$\begin{vmatrix} 0.4 & 0.6 & 0 \\ 0 & 0.5 & 0.5 \\ 0.3 & 0.3 & 0.4 \end{vmatrix}$$

If the chain starts in $X = 0$ at time $n = 0$, find

a) the transition matrix $\mathbf{P}(0, n)$ and the probability $B(n)$ that the chain is either in state 1 or in state 2 at time $n = 2$;

b) the first order distribution $\mathbf{\Pi}(n)$, and the second order distribution $\mathbf{\Pi}(n, n+1)$ (all terms) at time $n = \infty$;

c) the first order distribution $\mathbf{\Pi}(n)$ at time $n = 4$ knowing that the chain is in state 2 at time $n = 2$; the same, knowing that the chain is in the space subset $(1, 2)$ at time $n = 2$;

d) the first order distribution $\mathbf{\Pi}(n)$ at time $n + 2$ knowing that the chain is in the the space subset $(1, 2)$ at time n with $n = \infty$.

P.1.52 A Markov Chain presents the following one-step transition matrix at all times n.

$$\begin{vmatrix} 1/4 & 1/2 & 1/4 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{vmatrix}$$

If the chain starts with distribution $[0.5 \;\; 0.5 \;\; 0]$ at time $n = 0$, find

a) the distribution $\mathbf{\Pi}(2)$ and the asymptotic distribution;

b) the probability that the chain is in state 2 at time $n = 3$, and the probability that the chain is in state 1 at time $n = 2$ knowing that is in state 2 at time $n = 3$.

c) knowing that at time n the chain presents the asymptotic distribution, find the distribution at time $n - 1$; then find the distribution at time $n - 1$ when at time n the chain is in state 2.

P.1.53 A discrete-time random walk is modified in this way. Upon entering a state $i > 0$ the chain remains in the state for three time units, then leaves and with probability p and $1 - p$ goes to state $i + 1$ and $i - 1$ respectively. Leaving state 0, again after three time units, the chain reaches state 1 with probability 1.

a) Draw the state diagram and find the asymptotic distribution of the chain and the stability conditions on p;

b) the same as the above if the sojourn time is changed to a negative exponential variable with rate μ;

c) the same as point a) if sojourn time in state 0 is changed to one unit time.

P.1.54 A Markov Chain presents the following one-step transition matrix at all times n.

$$\begin{vmatrix} 0 & 1 & 0 \\ 0.4 & 0 & 0.6 \\ 0 & 1 & 0 \end{vmatrix}$$

If the chain starts in $X = 0$ at time $n = 0$, find

a) the first order distribution $\mathbf{\Pi}(n)$, and the transition matrix $\mathbf{P}(0, n)$ at times $n = 1, 2, 3, 4, \infty$;

b) the second order distribution $\mathbf{\Pi}(n-1, n)$ (all terms) at times $n = 3$ and $n = \infty$;

c) same as question a) starting at time $n = 0$ with the distribution $[0.5 \;\; 0.5 \;\; 0]$;

1.1 Problems' solutions- Chapter 1

P.1.1 Find the probability that k Poisson Arrivals at rate λ lie within an interval whose width is negative exponential RV of rate μ.

Solution. Thanks to the theorem of the total probability we have

$$P_k = \int_0^\infty \frac{(\lambda x)^k}{k!} e^{-\lambda x} \mu e^{-\mu x} dx$$

The integral is solved by parts, (is the integral of a Gamma function) or noting that it can be written as

$$P_k = \frac{\mu}{\lambda+\mu}\left(\frac{\lambda}{\lambda+\mu}\right)^k \int_0^\infty \frac{((\lambda+\mu)x)^k}{k!}(\lambda+\mu)e^{-(\lambda+\mu)x}dx = \frac{\mu}{\lambda+\mu}\left(\frac{\lambda}{\lambda+\mu}\right)^k$$

since the function within the integral is a pdf, namely the Erlang k +1, whose area is 1.

P.1.2 Two Poisson Arrival streams, of rate λ and μ respectively, merge on the same axis. Determine

a) The rate of the composite flow and probability that an arrival in the composite flow belongs to the first stream;

b) the probability that between two consecutive points of the second flow there are k of the first. (remember that $\int_0^\infty x^k e^{-x} dx = k!$)

Solution- a) The composite rate is the sum of the two, $\lambda + \mu$. The probability that an arrival is of the first type is a conditional probability, where the condition is the occurrence of an arrival. Denote $A = \{\text{arrival of the first type}\}$, $B = \{\text{arrival of the second type}\}$

$$p = P(A|A+B) = \frac{P(A(A+B))}{P(A+B)} = \frac{P(A)}{P(A+B)} = \frac{\lambda\Delta t}{\lambda\Delta t + \mu\Delta t} = \frac{\lambda}{\lambda+\mu}$$

b) The sought probability is obtained by the total probability theorem.

$$p_k = \int_0^\infty \frac{(\lambda x)^k}{k!} e^{-\lambda x}\mu e^{-\mu x} dx = \frac{\lambda^k \mu}{(\lambda+\mu)^{k+1}} \int_0^\infty \frac{((\lambda+\mu)x)^k}{k!}(\lambda+\mu)e^{-(\lambda+\mu)x}dx$$

We recognize that inside the integral operation there is a pdf, whose area is 1. Therefore we have

$$p_k = \frac{\lambda^k \mu}{(\lambda+\mu)^{k+1}}, \qquad k \geq 0 \qquad \text{(geometric distr.)}$$

Alternatively, by point a) each point has probability $\lambda/(\lambda+\mu)$ to be of the first type (success) independently of the other. So there must be k success and only one failure failure:

$$p_k = \frac{\mu}{\lambda+\mu}\left(\frac{\lambda}{\lambda+\mu}\right)^k$$

P.1.3 Two Poisson Arrival streams, A and B, of rate λ_A and λ_B respectively, merge on the same axis. Determine

a) The distribution of the number of arrivals in $[0;T]$;

b) the probability that an arrival in the composite flow belongs to the first stream;

c) the probability that an A arrival is followed by another A arrival;

d) If case c) occurs, find the pdf of the distance of such arrivals (a conditional pdf)

Solution

a)

$$P(N(0;T)=k)=\frac{(\lambda T)^k}{k!}e^{-\lambda T} \qquad k\geq 0$$

with $\lambda=\lambda_A+\lambda_B$.

b)

$$P_A=\frac{\lambda_A}{\lambda_A+\lambda_B}$$

c) The sought probability is the probability that the interarrival time of type A arrivals, X_a, is less than the interarrival time of type B arrivals, X_B:

$$P(X_A<X_B)=\frac{\lambda_A}{\lambda_A+\lambda_B}$$

Note that the result coincides with the result of point b). This is not incidental.

d) Bayes

$$f_{X_A}(x/X_A<X_B)=\frac{P(X_A<X_B/X_A=x)f_{X_A}(x)}{P(X_A<X_B)}=\frac{e^{-\lambda_B x}\lambda_A e^{-\lambda_A x}}{\dfrac{\lambda_A}{\lambda_A+\lambda_B}}=$$

$$f_{X_A}(x/X_A<X_B)=(\lambda_A+\lambda_B)e^{(\lambda_A+\lambda_B)x}$$

P.1.4 At a bus stop busses arrive according to a Poisson process of rate μ. Passengers arrive at the bus stop according to a Poisson process with rate λ. A bus arrives at the stop and finds zero passengers waiting.

a) find the pdf of the time elapsed since the last bus arrived.

b) and if the number of waiting is one?

Solution.

a)We use Bayes' theorem in the form

$$f_X(x/N(X)=0)=\frac{P(N(X)=0/X=x)f_X(x)}{\int P(N(X)=0/X=x)f_X(x)dx)}=\frac{e^{-\lambda x}\mu e^{-\mu x}}{\int_0^\infty e^{-\lambda x}\mu e^{-\mu x}dx}=$$

$$=(\lambda+\mu)e^{-(\lambda+\mu)x}$$

The result is a negative exponential pdf with parameter $\lambda + \mu$.

b) By repeating the argument with $N(x) = 1$ we have:

$$f_X(x/N(X)=1) = \frac{\lambda x e^{-\lambda x} \mu e^{-\mu x}}{\int_0^\infty \lambda x e^{-\lambda x} \mu e^{-\mu x} dx} = (\lambda+\mu)^2 x e^{-(\lambda+\mu)x}$$

i.e., the pdf is Erlang-2.

P.1.5 Consider Poisson arrival of rate λ. Someone tells us that within interval $[0;T]$ lie n arrivals. Find the the distribution of the number of arrivals M that lies in interval $[0;t], t \leq T$.

Solution.
We must find $P(M = k/N = n)$, whose definition is

$$P(M=k/N=n) = \frac{P(N=n; M=k)}{P(N=n)} = \frac{P(S=n-k; M=k)}{P(N=n)}$$

where we have denoted by S the number of arrivals in the second part of the interval. RVs M and S are statistically independent, as they relates to disjoint intervals and, therefore we have

$$P(M=k/N=n) = \frac{\frac{(\lambda(T-t))^{n-k}}{(n-k)!} e^{-\lambda(T-t)} \; \frac{(\lambda t)^k}{k!} e^{-\lambda t}}{\frac{(\lambda T)^n}{n!} e^{-\lambda T}} =$$

$$= \frac{n!}{k!(n-k)!} (\frac{T-t}{T})^{(n-k)} (\frac{t}{T})^k$$

Note that this is the same result we attain if n random points are distributed uniformly within $[0;T]$. Therefore, if we know that exactly n Poisson points lie in $[0;T]$, these behave exactly as n points uniformly taken within $[0;T]$.

P.1.6 Let us take the origin of time axis exactly at the time of a Poisson arrival, where the parameter is λ. Now take time τ. Find the pdf of Z, the distance from τ of the time instant where the last arrival before τ occurred.

Solution- If the origin does not coincide with a Poisson arrival, then the last arrival before t occurs at a distance X that is a negative exponential RV with parameter λ. When the origin coincides with a Poisson arrival, we have $Z = X$, if $X \leq \tau$, and $Z = \tau$ if $X > \tau$, which happens with probability $e^{-\lambda\tau}$. The solution is a mixed-type (continuous-discrete variable) pdf:

$$f_Z(z) = \lambda e^{-\lambda z} + e^{-\lambda\tau}\delta(z-\tau), \qquad 0 \leq z \leq \tau.$$

Alternatively, $P(Z > z),\ z < \tau$ is the probability we have no arrivals in z, i,e, interval from $\tau - z$ to τ, which happens with probability $P(Z > z) = e^{-\lambda\tau},\ z \leq \tau$. Then we have $P(Z > z) = 0,\ z > \tau$, i.e., a discontinuity in $z = \tau$. The derivative of $P(Z \leq z)$ provides the density.

P.1.7 Show that from the Markov property we also have

$$P(X_{n+2}, X_{n+1}/X_0, \dots X_n) = P(X_{n+2}, X_{n+1}/X_n)$$

Solution.
By the conditional probability formula, repeatedly applied, we have

$$P(X_{n+2}, X_{n+1}/X_0, \dots X_n) = P(X_{n+2}/X_0, \dots X_n, X_{n+1})P(X_{n+1}/X_0, \dots X_n) =$$

$$= P(X_{n+2}/X_n, X_{n+1})P(X_{n+1}/X_n) = P(X_{n+2}, X_{n+1}/X_n)$$

P.1.8 The binary MC X_n ($x_n = 0, 1$) presents the following one step transition matrices

$$\mathbf{P}(n, n+1) = \begin{vmatrix} 0.3 & 0.7 \\ 0.2 & 0.8 \end{vmatrix}$$

If the chain starts in $X = 0$ at time $n = 0$. Find

- the first order distribution at times $n = 1$ and $n = 2$;
- the joint second order distribution at times $1, 2$;
- the third order distribution at times $1, 2, 3$.

Solution
a)

$$\mathbf{\Pi}(1) = [0.3 \;\; 0.7], \qquad\qquad \mathbf{\Pi}(2) = [0.23 \;\; 0.77].$$

b) We make use of the expression

$$P(X_0, X_1, \dots, X_n, X_{n+1}) = P(X_0)P(X_1|X_0)P(X_2|X_1)\dots\dots P(X_n|X_{n-1}), P(X_{n+1}|X_n).$$

P.1.9 A Markov Chain presents the following one-step transition matrix at all times n.

$$\begin{vmatrix} 0.4 & 0.6 & 0 \\ 0 & 0.5 & 0.5 \\ 0.3 & 0.3 & 0.4 \end{vmatrix}$$

If the chain starts in $X = 0$ at time $n = 0$, find

a) the first order distribution $\mathbf{\Pi}(n)$, the transition matrix $\mathbf{P}(0, n)$, and the second order distribution $\mathbf{\Pi}(n, n+1)$ (all terms) all at time $n = 2$;

b) the probability $B(n)$ that the chain is either in state 1 or in state 2 at time $n = 2$; the first order distribution $\mathbf{\Pi}(n)$ at time $n = 4$ knowing that the chain is in state 2 at time $n = 2$; the same, knowing that the chain is in the space subset $(1, 2)$ at time $n = 2$;

Solution
a)

$$\mathbf{\Pi}(2) = [0.16 \;\; 0.54 \;\; 0.30], \qquad \mathbf{P}(0,2) = \begin{vmatrix} 0.16 & 0.54 & 0.30 \\ 0.15 & 0.40 & 0.45 \\ 0.24 & 0.45 & 0.31 \end{vmatrix}$$

$\mathbf{\Pi}(2,3)$ is obtained by multiplying the first row of $\mathbf{P}$ by $\pi_0(2)$, the second row of $\mathbf{P}$ by $\pi_1(2)$, and the third row of $\mathbf{P}$ by $\pi_2(2)$. We get

$$\mathbf{\Pi}(2,3) = \begin{vmatrix} 0.064 & 0.096 & 0 \\ 0 & 0.270 & 0.270 \\ 0.090 & 0.090 & 0.120 \end{vmatrix}$$

b)

$$B(2) = \pi_1(2) + \pi_2(2) = 0.84.$$

Knowing that the chain is in state 2 at time $n = 2$ means $\mathbf{\Pi}(2) = [0 \;\; 0 \;\; 1]$. Therefore

$$[0 \;\; 0 \;\; 1]\mathbf{P}(0,2) = [0.24 \;\; 0.45 \;\; 0.31].$$

Knowing that the chain is in the space subset $(1,2)$ at time $n = 2$ means that the distribution at time $n = 2$ is

$$\left[0 \;\; \frac{0.54}{0.84} \;\; \frac{0.30}{0.84}\right] = [0 \;\; 0.6429 \;\; 0.3571],$$

i.e., the conditional distribution the Chain has at time $n = 2$, starting in state 0 at time 0, conditional to $X(2) \in (1,2)$. The required distribution is

$$[0 \;\; 0.6429 \;\; 0.3571]\mathbf{P}(0,2) = [0.1821 \;\; 0.4179 \;\; 0.4000].$$

P.1.10 Two urns contain each N balls, N of which are blue and the others red, subdivided at random between the urns. At each instant, draw a ball from both urns and, with probability 1/2 exchange urn. The number of red balls in the first urn, $X(n)$ is Markov chain. Assuming $N = 2$, we write the one-step transition matrix and find the first order distribution at times $n = 1, 2, 3$ assuming that at the time $n = 0$ all the red balls are in the first urn.

Solution.
The one-step transition matrix is

$$\mathbf{P}(n, n+1) = \begin{vmatrix} 0.5 & 0.5 & 0 \\ 0.125 & 0.75 & 0.125 \\ 0 & 0.5 & 0.5 \end{vmatrix}$$

To get the first order distribution we use

$$\mathbf{\Pi}(n+1) = \mathbf{\Pi}(n)\mathbf{P}(n, n+1)$$

starting from $\mathbf{\Pi}(0) = [1; 0; 0]$. We have

$$\mathbf{\Pi}(1) = [0.5; 0.5; 0] \qquad \mathbf{\Pi}(2) = [0.3125; 0.625; 0.0625]$$

P.1.11 (The ruin problem) Two gamblers bet against each other, having probability 1/2 to win. Each time they bet one unit and the game ends when one of the two has not capital left. The capital of the first player is a Markov chain. In case the players start with capitals 1 and 2 respectively, write the one-step transition matrix and find the first order distribution at times $n = 1, 2, 3$.

Solution.
The state space is $0, 1, 2, 3$, and the corresponding one-step transition matrix is

$$\mathbf{P} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 1 \end{vmatrix}$$

To get the first order distribution we use

$$\mathbf{\Pi}(n+1) = \mathbf{\Pi}(n)\mathbf{P}(n, n+1)$$

starting from $\mathbf{\Pi}(0) = [0; 1; 0; 0]$. We have

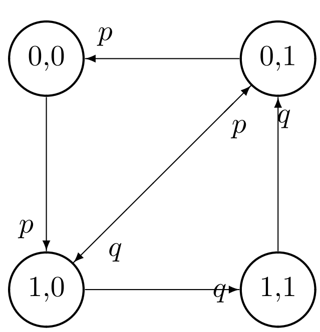
$$\mathbf{\Pi}(1) = [0.5 \;\; 0 \;\; 0.5 \;\; 0] \qquad \mathbf{\Pi}(2) = [0.5 \;\; 0.25 \;\; 0 \;\; 0.25]$$

P.1.12 Proces $Y(n)$ is defined by:

$$Y_{n+1} = Y_n \oplus Y_{n-1} \oplus X_{n+1}$$

where $\oplus$ is the binary summation and X_i is a sequence of binary $[0; 1]$ and independent RVs' with $P(X_i = 1) = p$. Show that $Y(n)$ is not (first order) Markov. The chain becomes Markov by re-defining states. Find the new state variable and draw the corresponding state diagram.

Solution.
From the very same formula that defines the process, the state at time $n + 1$ depends on the state at time n and on the state at time $n - 1$. Therefore, the process is second-order Markov. We can adopt the MC $Z_n = [Y_n, Y_{n-1}]$, and we obtain the following state diagram



P.1.13 Process X_n is a binary MC $[0; 1]$. Say whether process $Y(n)$:

$$Y_n = X_n \oplus X_{n-1}$$

is a MC.

Solution.
From relation

$$Y_{n-1} = X_{n-1} \oplus X_{n-2}$$

thanks to the fact that binary summation and difference coincide, we get

$$X_{n-1} = Y_{n-1} \oplus X_{n-2},$$

which, replaced in the original process provides

$$Y_n = X_n \oplus Y_{n-1} \oplus X_{n-2}.$$

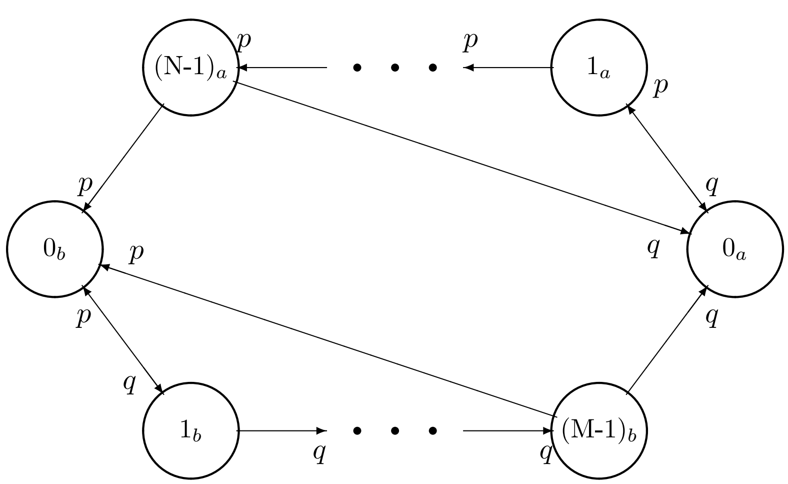
By repeating the steps above we see that we can write

$$Y_n = X_n \oplus Y_{n-1} \oplus Y_{n-2} \oplus Y_{n-3} \oplus \dots$$

Hence, even knowing Y_{n-1}, tha value of Y_n depends on the entire past history, and it can not be Markov of any order.

P.1.14 In a digital communication system symbols are continuously transmitted and p is the probability that there is a receiving error. The control system declares the system out of service ($X(t) = 0$) if N consecutive errors are discovered. On the other side, if the system is down, it is declared operational ($X(t) = 1$) after M correct and consecutive symbols are received. Show that the binary state $X(t)$ is not Markov, and define new states such that the new process $Y(t)$, derived by $X(t)$, is Markov. Finally, draw the state transition diagram for $Y(t)$.

Solution.
The operation of the system presents two distinct phases: in the former (index a) the system looks for N consecutive errors, whereas in the latter (index b) it looks for M consecutive correct symbols. The system can then be described by the Markov chain whose diagram of transitions is illustrated in the following figure, where we have used $q = 1 - p$,:



P.1.15 Given the following MCs, $X_{n+1} = X_n \oplus V_n$ and $Y_{n+1} = Y_n \oplus W_n$ Where $\oplus$ means binary summation, and the following binary and independent RVs V_n e W_n, with distribution respectively equal to p, $1-p$, and q, $1-q$. Check whether process $Z_n = X_n \oplus Y_n$, is markovian.

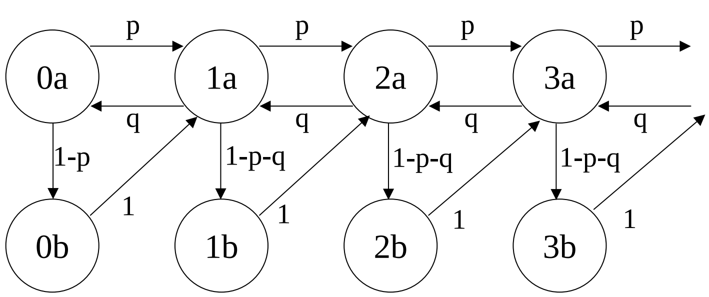
Solution. We have

$$Z_{n+1} = X_{n+1} \oplus Y_{n+1} = X_n \oplus V_n \oplus Y_n \oplus W_n = Z_n \oplus V_n \oplus W_n = Z_n \oplus T_n.$$

We see that future Z_{n+1} only depends on Z_n and, therefore, is markovian.

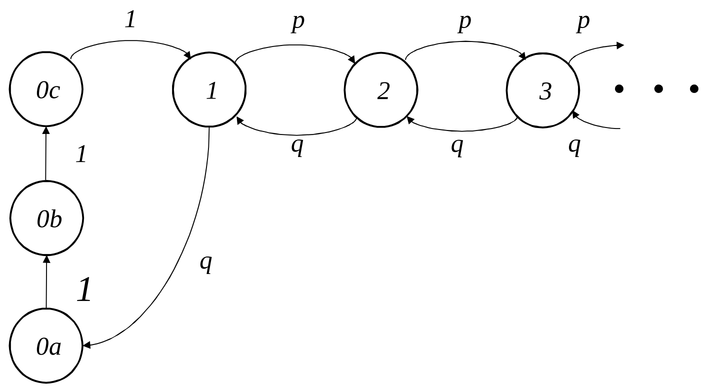
P.1.16 Let $X(n)$ be a random walk with a reflecting barrier in zero, with parameters p and q, $p+q<1$, where the residence period in each state can not exceed two time units. At the first time unit in state i transitions probabilities are p and q, as in the original walk. At the second time unit the only transition is toward state $i+1$ with probability one. Chain $X(n)$ is not Markov, but extending appropriately the state space, it is possible to define a Markov chain $Y(n)$ able to represent $X(n)$. Draw the state transition diagram for $Y(t)$.

Solution
The state diagram of $Y(n)$ is expanded by doubling the states of $X(n)$, where states i_a, represent the states i when entered coming from other states, and states i_b, represent the states i when otherwise. The diagram is as follows



P.1.17 Let $X(n)$ be a random walk with a reflecting barrier in zero, with parameters p and q, $p+q<1$, where we operate the following modifications. When the chain enters state zero, it remains there for two further time units (i.e., three time units in total), after wich it goes to state 1 with certainty. Chain $X(n)$ is not Markov, but extending appropriately the state space, it is possible to define a Markov chain $Y(n)$ able to represent $X(n)$. Draw the state transition diagram for $Y(n)$.

Solution
The state diagram of $Y(n)$ is as follows



P.1.18 Check whether distribution [2/9; 7/9] is a stationary distribution for the chain in Problem P.1.8

Solution.
We must check whether vector $\mathbf{\Pi}$, obeys to relation,

$$\mathbf{\Pi} = \mathbf{\Pi}\mathbf{P}(n, n+1),$$

which is true.

P.1.19 Check whether distribution [1/6; 4/6; 1/6] is a stationary distribution for the chain in Problem P.1.9

Solution.
As said above. The condition is verified.

P.1.20 Let $\mathbf{P}$ be the one-step transition matrix of the homogeneous MC $X(n)$. Take the chain $Y(n) = X(-n)$, shows it is a MC and write its one-step transition matrix $\mathbf{P}'$ (i.e., derive the new p'_{jk} as function of the old p_{jk} and ...).

P.1.21 The binary MC, X_n, ($x_n = 0, 1$), presents the following one step transition matrix

$$\mathbf{P}(n, n+1) = \begin{vmatrix} 1/4 & 3/4 \\ 1/2 & 1/2 \end{vmatrix}, \qquad \forall n.$$

If the chain starts in $X = 0$ at time $n = 0$, find

a) the first order distribution $\pi_j(n)$ (all terms) at times $n = 2$ and $n = \infty$;
b) the second order distribution $\pi_{jk}(n-1, n)$ (all terms) at times $n = 2$ and $n = \infty$;
c) the vector of the initial conditions $[\pi_0(0), \pi_1(0)]$, that makes $\pi_j(n)$ stationary at all $n > 0$.
d) the transition matrices $\mathbf{P}(0, 2)$ and $\mathbf{P}(0, n)$ with $n \to \infty$ (suggestion: note that $p_{jk}(0, \infty) = ...$).

Solution a) We have to evaluate

$$\mathbf{\Pi}(2) = \mathbf{\Pi}(0) \times \mathbf{P}^2 = [7/16; 9/16].$$

The second answer is clearly the asymptotic distribution:

$$\lim_{n\to\infty} \mathbf{\Pi}(n) = [2/5; 3/5].$$

b) Each element of the distribution can be evaluated as

$$\pi_{ij}(1,2) = \pi_i(1)p_{ij}.$$

The result is

$$\mathbf{\Pi}(1,2) = \begin{vmatrix} 1/16 & 3/16 \\ 6/16 & 6/16 \end{vmatrix}.$$

Again, each element of the distribution can be evaluated as

$$\lim_{n\to\infty} \pi_{i,j}(n-1,n) = \lim_{n\to\infty} \pi_i(n-1)p_{ij} = (\lim_{n\to\infty} \pi_i(n-1))p_{ij}.$$

The result is

$$\lim_{n\to\infty} \mathbf{\Pi}(n-1,n) = \begin{vmatrix} 2/20 & 6/20 \\ 6/20 & 6/20 \end{vmatrix}.$$

c) We must start with the stationary (asymptotic) distribution

$$\mathbf{\Pi}(0) = \mathbf{\Pi}(n) = [2/5; 3/5],$$

since this assure us that $\mathbf{\Pi}(1) = \mathbf{\Pi}(0)\mathbf{P} = \mathbf{\Pi}(0)$.

d)

$$\mathbf{P}(0,2) = \mathbf{P}^2 = \begin{vmatrix} 7/16 & 9/16 \\ 6/16 & 10/16 \end{vmatrix}.$$

For the second part, we must note that

$$\lim_{n\to\infty} p_{i,j}(0,n) = \pi_j,$$

since the chain is irreducible and the initial condition (state i) does not affect the asymptotic distribution. So we have

$$\lim_{n\to\infty} \mathbf{P}(0,n) = \begin{vmatrix} 2/5 & 3/5 \\ 2/5 & 3/5 \end{vmatrix}.$$

P.1.22 From a stack of N books customers take one at random, browse it and put it back at the top of the stack. The position on the stack $X(n)$ of a particular book is a MC.

a) Trace the state diagram of the MC

b) find the asymptotic probability that the book is found at the top of the stack.

Solution. a) Indicated with $x = 1$ position at the top of the stack, the non-null transition probabilities, except for the loop back transitions, are

$$p_{i,i+1} = \frac{N-i}{N} \qquad\qquad p_{i,1} = \frac{1}{N}$$

b)The balance at node i is

$$\pi_i \left(\frac{1}{N} + \frac{N-i}{N} \right) = \pi_{i-1} \frac{N-i+1}{N}$$

and therefore we have $\pi_i = \pi_{i-1}$ for any i. Finally $\pi_i = 1/N$ for all i.

P.1.23 Find the asymptotic distribution of a random walk with parameters p and q, having reflecting barriers in states $-N$ and N.

Solution.
We can find it in different ways. One considers the random walk between boundaries 0 and $2N$, and than shifts the index of states. We have

$$\pi_i = (p/q)^{i+N} \frac{1-p/q}{1-(p/q)^{2N+1}} \qquad\qquad -N \leq i \leq N$$

P.1.24 Consider a random walk with parameters p and q, having reflecting barriers in states 0 and 2:

a) find all the nine elements of the joint asymptotic distribution $\pi_{j,k}(n, n+1)$
b) find all the nine elements of the joint asymptotic distribution $\pi_{j,k}(n, n+2)$
c) find all the nine elements of the joint asymptotic distribution $\lim_{r\to\infty} \pi_{j,k}(n, n+r)$.

Solution.
a) We have

$$\pi_{j,k}(n, n+1) = \pi_j(n)p_{j,k}$$

where $\pi_j(n)$ is the asymptotic distribution of the first order. Details are omitted.

b) We have

$$\pi_{j,k}(n, n+2) = \pi_j(n)p_{j,k}(2)$$

where $\pi_j(n)$ is the asymptotic distribution of the first order and $p_{j,k}(2)$ is the transition probability in two steps. The matrix of transition probabilities in two steps is the square of matrix of transition probabilities in one step. Details are omitted.

c)

$$\lim_{r\to\infty} \pi_{j,k}(n, n+r) = \pi_j(n) \lim_{r\to\infty} p_{j,k}(r)$$

where $\pi_j(n)$ is the asymptotic distribution of the first order and $p_{j,k}(r)$ is the transition probability in r steps. Note that we have

$$\lim_{r\to\infty} p_{j,k}(r) = \lim_{r\to\infty} \pi_k(r) = \pi_k.$$

Finally

$$\lim_{r\to\infty} \pi_{j,k}(n, n+r) = \pi_j\pi_k$$

P.1.25 Find the asymptotic distribution of $Y(n)$ in problem P.1.12.

Solution- Solving the balance equations we find:

$$\pi_{0,0} = \pi_{0,1} = \pi_{1,0} = \pi_{1,1} = \frac{1}{4}$$

Then, we have :

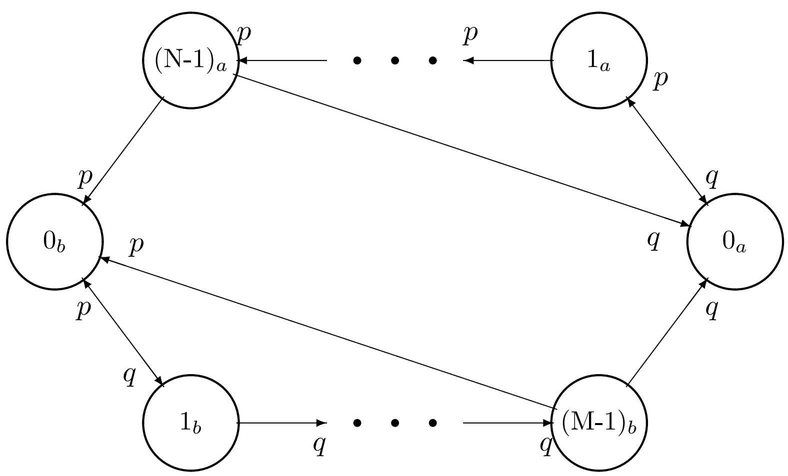
$$\lim_{n\to\infty} Pr[Y_n = 0] = \lim_{n\to\infty} Pr[Y_n = 0, Y_{n-1} = 0] + \lim_{n\to\infty} Pr[Y_n = 0, Y_{n-1} = 1] = \pi_{0,0} + \pi_{0,1} = \frac{1}{2}$$

and, similarly,:

$$\lim_{n\to\infty} Pr[Y_n = 0] = \pi_{1,0} + \pi_{1,1} = \frac{1}{2}$$

P.1.26 Find the asymptotic probabilities that the system in Problem P.1.14 is out of service (outage).

Solution- The state diagram is



The balance equations are:

$$\begin{cases} \pi_{i_a} = p\pi_{(i-1)_a} & i = 1, 2, \dots, N-1 \\ \pi_{i_b} = q\pi_{(i-1)_b} & i = 1, 2, \dots, M-1 \\ p\pi_{0_a} = q\sum_{j=1}^{N-1} \pi_{j_a} + q\pi_{(M-1)_b} \\ q\pi_{0_b} = p\sum_{j=1}^{M-1} \pi_{j_b} + p\pi_{(N-1)_a} \end{cases}$$

From the first two we get

$$\pi_{j_a} = p^j \pi_{0_a} \quad j = 1, 2, \dots, N-1$$
$$\pi_{j_b} = q^j \pi_{0_b} \quad j = 1, 2, \dots, M-1$$

and replacing into the fourth:

$$q^M \pi_{0_b} = p^N \pi_{0_a}$$

that is,

$$\pi_{0_a} = \frac{q^M}{p^N} \pi_{0_b}$$

Adding the congruence condition $\sum \pi_i = 1$ we get

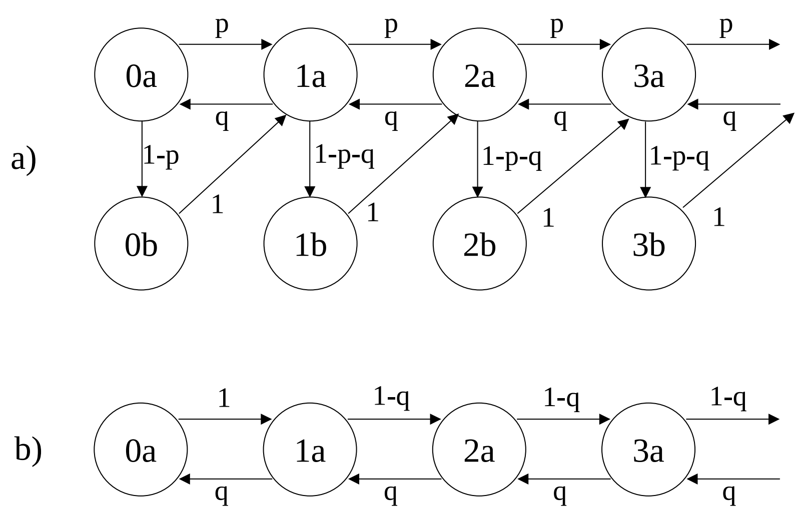
$$\pi_{0_b} = \frac{p^N}{q^{M-1}(1-p^N) + p^{N-1}(1-q^M)}$$

The outage probability is then:

$$Pr[\textbf{Outage}] = \sum_{j=0}^{M-1} \pi_{j_b} = \frac{1-q^M}{1-q} \pi_{0_b} = \frac{p^{N-1}(1-q^M)}{p^{N-1}(1-q^M) + q^{M-1}(1-p^N)}$$

P.1.27 Find the asymptotic distribution of the chain in problem P.1.16.

Solution-The state diagram for $Y(n)$ is shown in Figure a).



The stationary distribution can be obtained by writing the balance of fluxex across appropriate cuts. We can, however, take a shortcut. States ib presents only one input and one output, so for balance of fluxes nothing changes if flow into the state ib was sent directly to state $(i+1)a$. The result is shown in part b) of the figure.

The distribution of figure b) is the one of the generalized random walk

$$\pi_{ia} = \pi_{0a} \frac{1}{q} \left(\frac{1-q}{q} \right)^{i-1}; \qquad i \geq 1$$

which exists only if $1-q < q$, or $q > 0.5$.

The balance at nodes ib yields

$$\pi_{0b} = \pi_{0a}(1-p)$$

$$\pi_{ib} = \pi_{ia}(1-p-q) = \pi_{0a}\frac{1}{q}\left(\frac{1-q}{q}\right)^{i-1}(1-p-q); \qquad i>0$$

and the congruence condition

$$\sum \pi_{ia} + \sum \pi_{ib} = 1$$

provides π_{0a}.

c) The stationary distribution of $X(n)$ is then

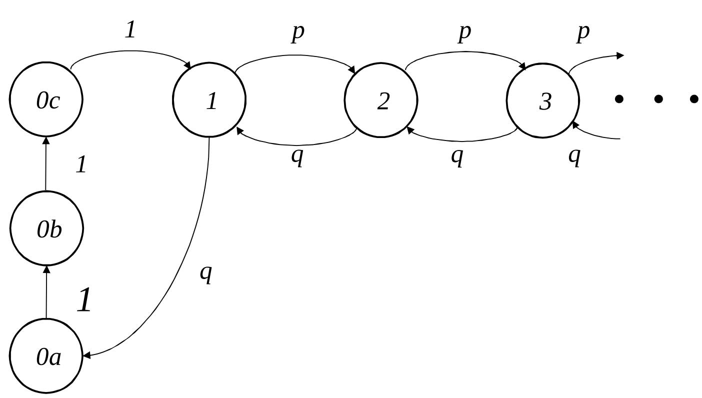
$$\pi_0 = \pi 0a + \pi_{0b} = \pi_{0a}(2-p)$$

$$\pi_i = \pi_{ia} + \pi_{ib} = \pi_{ia}(2-p-q) = \pi_{0a}\frac{1}{q}\left(\frac{1-q}{q}\right)^{i-1}(2-p-q); \qquad i \geq 1.$$

P.1.28 Find the asymptotic distribution of the chain in problem P.1.17.

Solution

The state diagram of $Y(n)$ is as follows



The cuts above state $0c$ shows that we have

$$\pi_i = \pi_{0c}\frac{1}{q}\left(\frac{p}{q}\right)^{i-1}; \qquad i \geq 1$$

$$\sum_{i\geq 1} \pi_i = \pi_{0c}\frac{1}{q}\sum_{i\geq 1}\left(\frac{p}{q}\right)^{i-1} = \pi_{0c}\frac{1}{q}\frac{1}{1-p/q} = \pi_{0c}\frac{1}{q-p}, \qquad p<q.$$

We also have

$$\pi_{0a} = \pi_{0b} = \pi_{0c},$$

and the congruence condition

$$\sum_{i\geq 1} \pi_i + 3\pi_{0c} = 1$$

provides π_{0c}.

If we are interested in the distribution π'_i of $X(n)$ we have

$$\pi'_i = \pi_i, \qquad i > 0, \qquad\qquad \pi'_0 = 3\pi_{0c}.$$

Note that the distribution π'_i of $X(n)$ is exactly the one we have when state 0 is markovian with transition probability toward state one equal to 1/3.

P.1.29 A discrete-time MC has the following transition matrix over states $0, 1, 2, 3, 4$:

$$\mathbf{P} = \begin{vmatrix} 0.2 & 0.8 & 0 & 0 & 0 \\ 0.4 & 0.6 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.8 & 0.2 \\ 0 & 0 & 0 & 0.3 & 0.7 \end{vmatrix}$$

Find

a) the distribution vector at time $n = 2$ in the two cases when it starts at time $n = 0$ in states $X = 0$ and $X = 4$ respectively;
b) the asymptotic distributions in the two cases in a)
c) verify whether the vector $[1/6; 2/6; 0; 3/10; 2/10]$ represents an asymptotic (stationary) distribution of the chain;

Solution
a) With the usual procedure we find, starting from 0 and 4 respectively:

$$\mathbf{\Pi}(2) = \mathbf{\Pi}(0)\mathbf{P}^2 = [0.36, 0.64, 0, 0, 0]$$

$$\mathbf{\Pi}(2) = \mathbf{\Pi}(4)\mathbf{P}^2 = [0, 0, 0, 0.45, 0.55]$$

b) By observing the state space of the chain we see that subsets $0, 1$ and $3, 4$ are irreducible. This means that once you start in each of them you can not leave. In other words, if you start in the former (state 0) you have a binary chain with asymptotic distribution $[1/3, 2/3]$, while if you start in the latter (state 4) you have a binary chain with asymptotic distribution $[3/5, 2/5]$. This leads to the following answer: , starting from 0 and 4 respectively we have:

$$\mathbf{\Pi} = [1/3, 2/3, 0, 0, 0]$$

$$\mathbf{\Pi} = [0, 0, 0, 3/5, 2/5]$$

c) We must verify whether the suggested distribution satisfies the balance equations

$$[1/6; 2/6; 0; 3/10; 2/10] = [1/6; 2/6; 0; 3/10; 2/10]\mathbf{P}.$$

The answer is Yes. In fact, the suggested distribution coincides with the asymptotic distribution when starting at time 0 in state 2.

P.1.30 A binary MC X_n has the following transition probabilities

$$p_{01}(n, n+1) = p_1, \qquad p_{10}(n, n+1) = q_1, \qquad n = 2k + 1$$

i.e., when n is odd, and

$$p_{01}(n, n+1) = p_2, \qquad p_{10}(n, n+1) = q_2, \qquad n = 2k$$

i.e., when n is even. Find the asymptotic distributions in even instants, in odd instants, and in any instant.

Solution.
The MC is non-homogeneous, i.e., the transition probability is time dependent: in this case it has a periodic behavior which assumes two distinct values in even and odd instants. The matrix of transitions at even instants is

$$\mathbf{P}^{e\to o} = \begin{bmatrix} 1-p_2 & p_2 \\ q_2 & 1-q_2 \end{bmatrix}$$

and the one at odd instants is

$$\mathbf{P}^{o\to e} = \begin{bmatrix} 1-p_1 & p_1 \\ q_1 & 1-q_1 \end{bmatrix}$$

The two steps transition matrix between even instants is

$$\mathbf{P}^{e\to e} = \mathbf{P}^{e\to o}\mathbf{P}^{o\to e},$$

does not depend on time, so that the chain of sole even instants is homogeneous. The balance equations of this chain provides the asymptotic distributions in even instants

$$\pi_0^e = \frac{q_2(1-p_1)+q_1(1-q_2)}{q_1(1-p_2)+q_2(1-q_1)+p_2(1-p_1)+p_1(1-q_2)}$$

$$\pi_1^e = 1-\pi_0^e$$

Similarly, for odd instants, we have

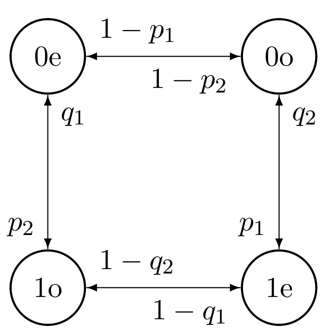
$$\mathbf{P}^{o\to o} = \mathbf{P}^{o\to e}\mathbf{P}^{e\to o}$$

the solution is the same of even instants but exchanging indexes 1 and 2.

As for the asymptotic distribution at any instant, we use the total probability formula (or the generalized asymptotic distribution) and get

$$\pi_0 = \frac{1}{2}(\pi_0^e+\pi_0^o)$$

Another way to arrive at the same solution consists in taking the memory into a new state definition and get in this way an homogeneous chain. More precisely, we consider the set of states $[0e, 1e, 0o, 1o]$ with obvious meaning of symbols, and we have the following transition diagram



The asymptotic probabilities of this chain are related to the preceding ones in the following way

$$
\begin{aligned}
\pi_0^o &= \frac{\pi_{0o}}{\pi_{0o}+\pi_{1o}} = 2\pi_{0o} \\
\pi_1^o &= \frac{\pi_{1o}}{\pi_{0o}+\pi_{1o}} = 2\pi_{1o} \\
\pi_0^e &= \frac{\pi_{0e}}{\pi_{0e}+\pi_{1e}} = 2\pi_{0e} \\
\pi_1^e &= \frac{\pi_{1e}}{\pi_{0e}+\pi_{1e}} = 2\pi_{1e} \\
\pi_0 &= \pi_{0o}+\pi_{0e} \\
\pi_1 &= \pi_{1o}+\pi_{1e}
\end{aligned}
$$

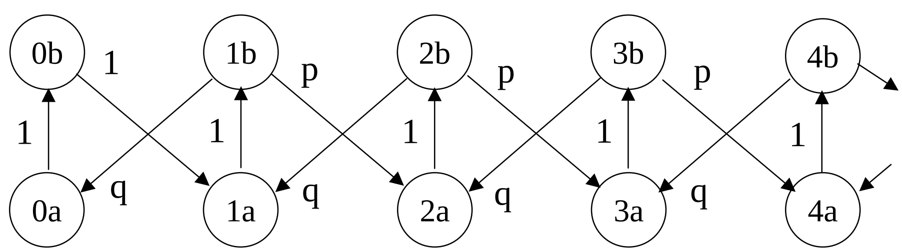
P.1.31 Let $X(n)$ be a random walk between states 0 and N where at each state $i > 0$ the process stops for the next step and and the second step goes up or down with probability respectively equal to p and $1-p$. For $i = 0$ in the process stops for a further step and then switches to $i = 1$ with certainty.

a) $X(n)$ is not an MC, but including some memory we can build a MC $Y(n)$. Draw state diagram of that chain;

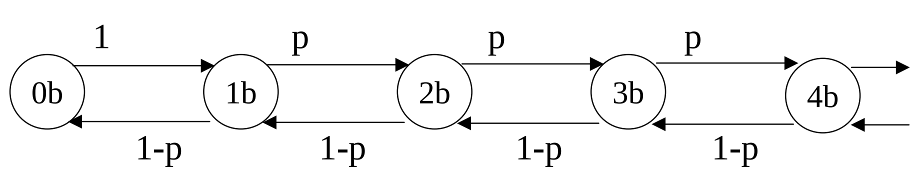
b) Find the asymptotic distribution of $X(n)$ if $N \to \infty$.

Solution

a) The state diagram is shown in the figure below with $q = 1-p$.



b) The balance at node b shows $\pi_{ia} = \pi_{ib}$. In this way the equations reduce to those of the chain in the figure below



whose solution is well known:

$$\pi_{ib} = \pi_{0b} \frac{p^{i-1}}{(1-p)^i}, \qquad i \geq 1,$$

and exists for $1 - p > p$, or $p < 0.5$. Then we have

$$\pi_i = \pi_{ia} + \pi_{ib} = \pi_0 \frac{p^{i-1}}{(1-p)^i}, \qquad i \geq 1,$$

$$\pi_0 = \frac{1-2p}{2-2p}.$$

P.1.32 A non-homogeneous MC is such that the matrix of transitions in a step $\mathbf{P}(n, n+1)$ is periodic with period equal to 3 starting from steps for $n = 0, 1, 2$, and taking the following values

$$\begin{vmatrix} 0.4 & 0.6 & 0 \\ 0.4 & 0 & 0.6 \\ 0 & 0.4 & 0.6 \end{vmatrix} \qquad \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} \qquad \begin{vmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix}$$

Staring at $n = 0$ in state 0 find

a) the distribution at times 1, 2 e 3.
b) the asymptotic distribution at times $3n$, $3n + 1$ e $3n + 2$, $n = 0, 1, 2, \ldots$;
c) the asymptotic distribution (i.e., in a RIP).

Solution
a)

$$\mathbf{\Pi}(1) = \mathbf{\Pi}(0)P(0) = [0.4; 0.6; 0]$$

$$\mathbf{\Pi}(2) = \mathbf{\Pi}(1)P(1) = [0.4; 0.6; 0]$$

$$\mathbf{\Pi}(3) = \mathbf{\Pi}(2)P(2) = [1; 0; 0]$$

b)

$$\mathbf{\Pi}(3n) = \mathbf{\Pi}(0) = [1; 0; 0]$$

$$\mathbf{\Pi}(3n+1) = \mathbf{\Pi}(1) = [0.4; 0.6; 0]$$

$$\mathbf{\Pi}(3n+2) = \mathbf{\Pi}(2) = [0.4; 0.6; 0]$$

c)

$$\frac{1}{3}\mathbf{\Pi}(3n) + \frac{1}{3}\mathbf{\Pi}(3n+1) + \frac{1}{3}\mathbf{\Pi}(3n+2) = [0.6; 0.4; 0]$$

P.1.33 (3.42) A bidimensional random walk in the discrete time moves on step in one of the four directions North, East, South, West, with probability respectively equal to p, p, q, q. Axis are reflective barriers so that the walk only occurs in the first quadrant (X and Y non-negative integers). Denoted by (i, j) the state, i.e., the position in the walk:

a) Check whether a distribution of the form $\pi_{i,j} = K(p/q)^{i+j}$ satisfies the balance equations, and if so find K.

b) Find the distribution of the walk on the horizontal line, i.e., The marginal π_i;

c) Same as a) and b) when the walk is limited by the reflecting barrier $X + Y = 2$.

Solution. a) The propose solution satisfies the balance equations; furthermore we have

$$\pi_{ij} = (1 - p/q)^2 (p/q)^{i+j}, \quad \forall i, j$$

b) The marginal distribution is

$$\pi_i = (1 - p/q)(p/q)^i, \quad \forall i$$

In particular we verify that

$$\pi_{ij} = \pi_i \pi_j \quad \forall i, j$$

so that the components are statistically independent. c) The general solution is the same as in a), because the dropped states do not alter the flux equilibrium. This means that the general solution is the same

$$\pi_{ij} = K(p/q)^{i+j}, \quad i + j \leq 2$$

with

$$K = \frac{1}{1 + 2(p/q) + 3(p/q)^2}.$$

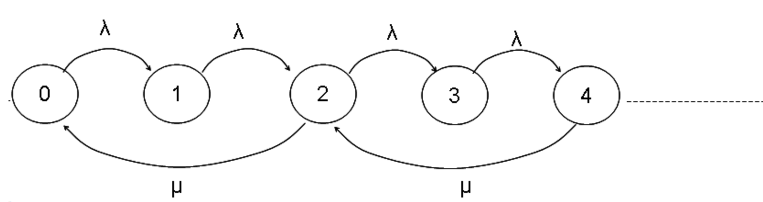
P.1.34 (4.1) The rate matrix $\mathbf{Q}$ of a continuous-time MC presents the following non-null elements outside the the main diagonal:

$$q_{i,i+1} = \lambda \qquad i = 0, 1, 2, 3, \ldots$$

$$q_{i,i-2} = \mu \qquad i = 2, 4, 6, \ldots$$

Find the equilibrium distribution π_i $\forall$ i. Is this chain reversible? Draw some sample paths.

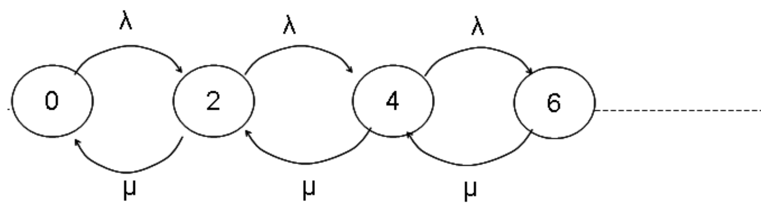
Solution.
The state transition diagram is reported in the following figure



In odd states the fluxes in and out are equal. Then we have

$$\pi_i = \pi_{i-1} \qquad\qquad i = 1, 3, 5, 7, \ldots$$

The solution for the even states can be obtained also solving the following chain:



and we get

$$\pi_i = \pi_0 \left(\frac{\lambda}{\mu}\right)^{i/2} \qquad\qquad i = 2, 4, 6, \ldots$$

Using $\sum_i \pi_i$ we get

$$\pi_0 = \frac{1}{2\sum_{i=0}^{\infty}(\frac{\lambda}{\mu})^i} = \frac{1}{2}\left(1 - \frac{\lambda}{\mu}\right).$$

P.1.35 (4.7) A MC is composed of two states, 0 and 1, where 1 is an absorbing state. Find the transient distribution $\pi_0(t)$ e $\pi_1(t)$ assuming that the chain starts in state 0.

Solution. Since state 1 is an absorbing state the only possible transition is $0 \to 1$. $\pi_0(t)$ is the probability that the transition has not occurred yet. If X denotes the time spent in state 0, negative exponential with rate, say, μ, we have

$$\pi_0(t) = P(X > t) = 1 - F_X(t) = e^{-\mu t} \qquad\qquad \pi_1(t) = 1 - \pi_0(t) = 1 - e^{-\mu t} \qquad (1.1)$$

P.1.36 (4.9) A process $X(t)$ leaves state 0 with rate μ and proceeds toward the states $1, 2, 3, 4$ successively, and stay in each of them for a period of time of length $i, i = 1, 2, 3, 4$. After leaving state 4 it returns to state 0, and the cycle re-starts again. Derive, explaining, the asymptotic distribution.

Solution. The prpocess is periodic, amd then is regenerative. The average cycle time is

$$E[C] = 1/\mu + 1 + 2 + 3 + 4 = 1/\mu + 10$$

and the distribution, because of the limit theorem on regenerative processes ($\pi_i = E[Y_i]/E[C]$) is

$$\pi_0 = \frac{1}{1 + 10\mu} \qquad \pi_i = \frac{i\mu}{1 + 10\mu} \quad i > 0$$

P.1.37 (4.16) A continuous-time binary chain $X(t)$ remains in state 1 a period of time τ and 2τ alternatively, while it still remains a period of time 2τ in state 0.

a) Find the asymptotic distribution;

b) find the average (asymptotic) rates q_{10} e q_{01}.

Solution.
a) $X(t)$ is a regenerative process with periodic cycles of length $C = 7\tau$. The distribution ($\pi_i = E[Y_i]/E[C]$) is

$$\pi_1 = \frac{3\tau}{7\tau} = 3/7 \qquad\qquad \pi_0 = \frac{4\tau}{7\tau} = 4/7$$

.

b) We have two ways. First, $q_{10}\Delta t$ is the conditional probability of transition $1 \to 0$ within $[t; t + \Delta t]$ where t is an asymptotic point, i.e., a Random Inspection Point. The probability that a RIP, taken while the chain is in state 1, lies within such Δt is

$$q^*_{10}\Delta t = \frac{2\Delta t}{3\tau}$$

In a similar way we have

$$q^*_{01}\Delta t = \frac{2\Delta t}{4\tau}$$

The second way is to evaluate the average sojourn time, since the rate coincides with its reverse. The sojourn time in 1 is alternatively τ and 2τ, with average 1.5τ, and its reverse is $(2/3\tau)$.

Note that the found probability and rates must still obey to the balance of fluxes

$$\pi_1 q_{10} = \pi_0 q_{01}$$

although the chain is not markovian.

P.1.38 Consider a time continuous random walk on the positive axis and reflecting barrier at the origin, with rates λ and μ. Assume now that when it enters state 0 it remains there for a constant time T and then goes to state 1. Find the asymptotic distribution.

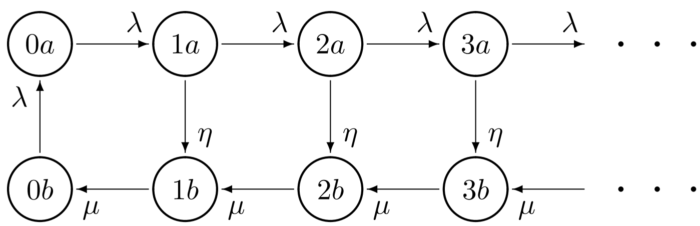
Solution. The modifications make the chain a semi-markov chain. According to what exposed in the notes, the distribution of the semi-markov can be attained as with Markov Chains replacing rates with the inverse of the average sojourn times. This makes the birth rate of state 0 equal to $1/T$, all other rates being unchanged, which provides

$$\pi'_i = \pi'_0 \frac{1}{\mu T}\left(\frac{\lambda}{\mu}\right)^{i-1} = \pi'_0 \frac{1}{\lambda T}\left(\frac{\lambda}{\mu}\right)^{i}, \quad i \geq 1,$$

$$\pi'_0 = \frac{\mu T}{\mu T + 1/(1 - \lambda/\mu)},$$

which turns out, no surprise, equal to case a).

P.1.39 (4.3) - A time continuous MC presents the following state diagram:



Find the asymptotic distribution π_{ix}, and the conditions on the parameters such that the asymptotic distribution exists. Check whether the distribution remains the same (up to a constant factor) when the chain is truncated to state N (ie, flux out of Na upward, and flux into Nb downward are canceled.

Solution a) By the balance of fluxes at nodes a we get

$$\pi_{ia} = \pi_{0a} \left(\frac{\lambda}{\lambda + \eta} \right)^i, \qquad i \geq 0.$$

By the balance of fluxes at a vertical cut we have

$$\pi_{ib} = \pi_{0a} \left(\frac{\lambda}{\lambda + \eta} \right)^{i-1} \frac{\lambda}{\mu} \qquad i \geq 1$$

$$\pi_{0b} = \pi_{0a}$$

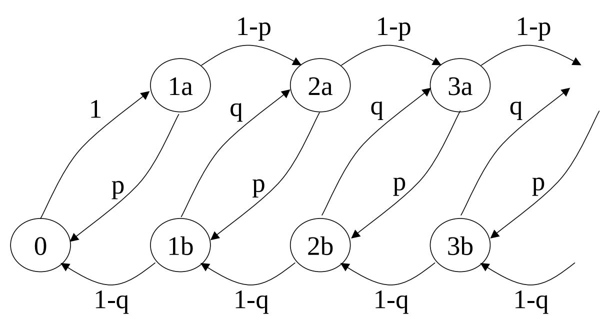
b) In order the solution to converge we must have $\eta > 0$.

c) The truncated chain is not balanced by the solution above.

P.1.40 A random walk process $N(n)$ is such that it moves upward and downward according to a pre-set direction. At each step the direction of the step is changed, before leaving the state, in this way: if the direction was upward it is changed with probability p; if the direction was downward it is changed with probability; q. Upon entering the origin, the direction is immediately reversed. Process $N(n)$ is not markovian, but we can derive its distribution defining a new process $Y(n)$ that is markovian.

a) draw the state diagram of process $Y(n)$;
b) find the asymptotic distribution of $N(n)$ together with the conditions on the parameters such that the asymptotic distribution exists.

Solution a)



b) From a diagonal cut we have

$$\pi_{ia}(1-p) = \pi_{ib}(1-q), \qquad i \geq 1,$$

and fom balance to b nodes

$$\pi_{ib} = \pi_{i+1,b}(1-q) + \pi_{i+1,a}p, \qquad i \geq 0.$$

By substituting we get

$$\pi_{i+1,b} = \pi_{ib}\frac{1-p}{1-q}, \qquad i \geq 0,$$

which yields

$$\pi_{ib} = \pi_0 \left(\frac{1-p}{1-q}\right)^i \qquad i \geq 0,$$

$$\pi_{ia} = \pi_0 \left(\frac{1-p}{1-q}\right)^{i-1} \qquad i \geq 1,$$

$$\pi_0 = \frac{p-q}{2(1-q)},$$

where the convergence of the solution exists for $p > q$. The distribution of the original chain is then

$$\pi_i = \pi_{ia} + \pi_{ib} \quad i \geq 1$$

P.1.41 A Markov Chain presents the following one-step transition matrix at all times n.

$$\begin{vmatrix} 0.4 & 0.6 & 0 \\ 0.4 & 0 & 0.6 \\ 0 & 0.4 & 0.6 \end{vmatrix}$$

It starts at time 0 in state 0 and at all time instants $n = 3k$, $k = 1, 2, 3, \ldots$, is forced to assume value 0. Find

a) the distribution at times 1, 2 e 3.
b) the asymptotic distribution ($n \to \infty$) at times $3n$, $3n+1$ e $3n+2$, $n = 0, 1, 2, \ldots$;
c) the asymptotic distribution (at time n, $n \to \infty$) (i.e., in a RIP).

Solution
a) With the usual procedure we find,

$$\mathbf{\Pi}(1) = \mathbf{\Pi}(0)\mathbf{P} = [0.4, 0.6, 0]$$

$$\mathbf{\Pi}(2) = \mathbf{\Pi}(1)\mathbf{P} = [0.4, 0.24, 0.36]$$

$$\mathbf{\Pi}(3) = \mathbf{\Pi}(0) = [1, 0, 0,]$$

b)

$$\mathbf{\Pi}(3n) = \mathbf{\Pi}(0)$$

$$\mathbf{\Pi}(3n+1) = \mathbf{\Pi}(1)$$

$$\mathbf{\Pi}(3n+2) = \mathbf{\Pi}(2)$$

c)

$$\mathbf{\Pi} = (1/3)(\mathbf{\Pi}(3n) + \mathbf{\Pi}(3n+1) + \mathbf{\Pi}(3n+2)) = [0.6, 0.28, 0.12]$$

P.1.42 Three cards are distributed in positions 0, 1, 2. At each step, the pair of positions 0 and 1 is selected with probability 1/3 and the cards exchange positions. Otherwise, the pair 1 and 2 is selected and the cards exchanged. If an Ace is placed in position 0, what is the distribution of its position after two steps? and after an infinite number of steps? What if, at each step, all pairs are selected with the same probability and the cards are exchanged?

Solution The position of the Ace is a homogeneous, irreducible Markov chain whose transition matrix is

$$P = \begin{vmatrix} 2/3 & 1/3 & 0 \\ 1/3 & 0 & 2/3 \\ 0 & 2/3 & 1/3 \end{vmatrix}$$

We have

$$\mathbf{\Pi}(0) = [1, 0, 0]$$

$$\mathbf{\Pi}(2) = \mathbf{\Pi}(0)\mathbf{P}^2 = [5/9, 2/9, 2/9]$$

Asymptotically we must solve the balance equations to get

$$\mathbf{\Pi} = [1/3, 1/3, 1/3]$$

The latter can be also inferred by the fact that the matrix has columns that sum to one.

When all pairs are selected with the same probability the matrix becomes

$$P = \begin{vmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{vmatrix}$$

From this we see that at each step we have distribution again equal to $[1/3, 1/3, 1/3]$.

P.1.43 A binary MC X_n has the following transition probabilities

$$p_{01}(n, n+1) = 1/2, \qquad p_{10}(n, n+1) = 1/2, \qquad n = 3k$$

$$p_{01}(n, n+1) = 1/4, \qquad p_{10}(n, n+1) = 3/4, \qquad n = 3k+1$$

$$p_{01}(n, n+1) = 1, \qquad p_{10}(n, n+1) = 1, \qquad n = 3k+2$$

Find

a) the distribution at times 1, 2 and 3, starting in state 0 at time 0.
b) the asymptotic distribution ($n \to \infty$) at times $3n$, $3n+1$ e $3n+2$, $n = 0, 1, 2, \dots$;
c) the asymptotic distribution (at time n, $n \to \infty$) (i.e., in a RIP).

Solution a), b) The MC is non-homogeneous, i.e., the transition probability is time dependent. In this case a simple inspection at times $1, 2, 3, 4, \dots$ shows that we have

$$\mathbf{\Pi}(3n+1) = [1/2; 1/2], \qquad \mathbf{\Pi}(3n+2) = [3/4; 1/4], \qquad \mathbf{\Pi}(3n+3) = [1/4; 3/4].$$

The above also provide the asymptotic distributions in the required instants.

As for the asymptotic distribution at any instant, we use the total probability formula and get

$$\pi_0 = \frac{1}{3}(\pi(3n+1) + \pi(3n+2) + \pi(3n+3)) = [1/2; 1/2].$$

P.1.44 A random walk between zero and infinite has parameters p and q, $p < q$. Find

a) the asymptotic joint probability $\pi_{j,k}(n, n+1)$ (any j, k);
b) the asymptotic joint probability $\pi_{j,k}(n, n+\infty)$ (any j, k);
c) the asymptotic probability of state j when arriving from state $j-1$, and the one when arriving from state $j+1$.

Solution
The asymptotic distribution at time n is

$$\pi_j = (1 - p/q)(p/q)^j, j \geq 0$$

a) At time $n+1$ we have a steps $+1, 0, -1$ with probabilities $p, 1-p-q, q$ respectively. Therefore we have

$$\pi_{j,j+1}(n, n+1) = p(1 - p/q)(p/q)^j$$
$$\pi_{j,j}(n, n+1) = (1 - p - q)(1 - p/q)(p/q)^j$$
$$\pi_{j,j-1}(n, n+1) = q(1 - p/q)(p/q)^j$$
$$\pi_{j,k}(n, n+1) = 0, \text{ for others values of } k,$$

and boundary conditions.

b) The state in two points very far away (infinite) are statistically independent. Therefore, We have

$$\pi_{j,k}(n, n+\infty) = \pi_j \pi_k$$

c) The asymptotic probability of state j when arriving from state $j-1$, is the conditional

$$\frac{\pi_{j-1,j}(n, n+1)}{\pi_{j-1}(n)} = \frac{p(1 - p/q)(p/q)^{j-1}}{(1 - p/q)(p/q)^{j-1}} = p$$

and when arriving from state $j+1$,

$$\frac{\pi_{j+1,j}(n, n+1)}{\pi_{j+1}(n)} = \frac{q(1 - p/q)(p/q)^{j+1}}{(1 - p/q)(p/q)^{j+1}} = q$$

P.1.45 Two discrete-time MC's have the following transition matrices

$$\mathbf{P}_1 = \begin{vmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{vmatrix} \qquad \mathbf{P}_2 = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}$$

A new chain starts at time $n = 0$ in state 0 using the first matrix. It uses the same matrix at $n = 1$; then at time $n = 2$ uses the second matrix, and switches back to the first matrix at time $n = 3$,and then repeats periodically; Find

a) the distribution vector at times $n = 2, 3, 4$, the transition matrices $\mathbf{P}(0, 2)$ and $\mathbf{P}(0, 3)$;
b) $\lim_{n\to\infty} P_1^n$;
c) the asymptotic distribution at times $n = 3k$, $k = 0, 1, 2, \ldots$ (note that it can be derived as a limit from $\mathbf{P}(0, 3)$.
d) the asymptotic distribution;

Solution
a)

$$\mathbf{\Pi}(2) = [0.48 \;\; 0.52] \qquad \mathbf{\Pi}(3) = [0.52 \;\; 0.48] \qquad \mathbf{\Pi}(4) = [0.4560 \;\; 0.5440]$$

$$\mathbf{P}(0, 2) = \begin{vmatrix} 0.48 & 0.52 \\ 0.39 & 0.61 \end{vmatrix} \qquad \mathbf{P}(0, 3) = \begin{vmatrix} 0.52 & 0.48 \\ 0.61 & 0.39 \end{vmatrix}$$

b) $\lim_{n\to\infty} \mathbf{P}_1^n$, represents the asymptotic behavior of the chain whose matrix is $\mathbf{P}_1$. Since $\mathbf{P}_1$ is irreducible and finite, the asymptotic matrix presents all rows equal to the asymptotic distribution of $\mathbf{P}_1$, i.e.,

$$\lim_{n\to\infty} \mathbf{P}_1^n = \begin{vmatrix} 3/7 & 4/7 \\ 3/7 & 4/7 \end{vmatrix}$$

c) The sought distribution is the asymptotic distribution of the chain whose matrix is $\mathbf{P}(0, 3)$, which, for result a) is $\mathbf{\Pi}' = [0.5596 \;\; 0.4404]$.

d) the asymptotic distribution at times $n = 3k + 1$, is $\mathbf{\Pi}'\mathbf{P}_1 = [0.4679 \;\; 0.5321]$, while the asymptotic distribution at times $n = 3k + 2$, is $\mathbf{\Pi}'\mathbf{P}_1\mathbf{P}_2 = [0.5321 \;\; 0.4679]$. Hence, the asymptotic distribution (at all times) is $\mathbf{\Pi} = [0.5199 \;\; 0.4801]$.

P.1.46 A Markov Chain presents the following one-step transition matrix at all times n.

$$\mathbf{P} = \begin{vmatrix} 0 & 0.6 & 0.4 \\ 0.4 & 0.6 & 0 \\ 0.6 & 0 & 0.4 \end{vmatrix}$$

If the chain starts in $X = 0$ at time $n = 0$, find

a) the first order distribution $\mathbf{\Pi}(n)$, and the transition matrix $\mathbf{P}(0, n)$ at time $n = 2$;
b) the first order distribution $\mathbf{\Pi}(n)$, and the transition matrix $\mathbf{P}(0, n)$ at time $n = \infty$;
c) the second order distribution $\mathbf{\Pi}(n - 1, n)$ (all terms) at times $n = 2$ and $n = \infty$;

d) the vector of the initial distribution $\mathbf{\Pi}(0)$ that makes the chain stationary at all $n \geq 0$.

Solution

a)

$$\mathbf{P}(0,2) = \begin{vmatrix} 0.48 & 0.36 & 0.16 \\ 0.24 & 0.6 & 0.16 \\ 0.24 & 0.36 & 0.4 \end{vmatrix} \qquad \mathbf{\Pi}(n) = [0.48 \;\; 0.36 \;\; 0.16]$$

b)

$$\mathbf{P}(0,\infty) = \begin{vmatrix} 6/19 & 9/19 & 4/19 \\ 6/19 & 9/19 & 4/19 \\ 6/19 & 9/19 & 4/19 \end{vmatrix} \qquad \mathbf{\Pi}(\infty) = [6/19 \;\; 9/19 \;\; 4/19]$$

c) the elements of $\mathbf{\Pi}(n-1,n)$ at time $n=2$

$$\mathbf{\Pi}(1,2) = \begin{vmatrix} 0 & 0 & 0 \\ 0.24 & 0.36 & 0 \\ 0.24 & 0 & 0.16 \end{vmatrix} \qquad \mathbf{\Pi}(\infty,\infty+1) = \begin{vmatrix} 0 & 36/190 & 24/190 \\ 36/190 & 54/190 & 0 \\ 24/190 & 0 & 16/190 \end{vmatrix}$$

d)

$$\mathbf{\Pi}(0) = [6/19 \;\; 9/19 \;\; 4/19]$$

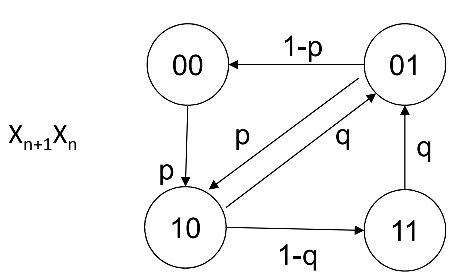
P.1.47 Given a two-state $(0,1)$ discrete-time Markov Chain $X(n),\ n = 0,1,\ldots$ with transition probabilities p and q, derive chain $Y(n),\ n = 1,2,\ldots$ such that $Y(n+1) = X(n) \oplus X(n+1)$, with $\oplus$ meaning binary summation. Find

a) the first order distribution $\mathbf{\Pi}(n)$ of $Y(n)$ at time $n=2$ starting at time $n=0$ in with $X(0)=0$ and $Y(0)=0$.

b) the asymptotic distribution $\mathbf{\Pi}$ of $Y(n)$.

c) the asymptotic second order distribution $\mathbf{\Pi}(n,n+1)$, i.e., $n=\infty$.

Solution - The state of chain $Y(n+1)$ (not Markov) depends on both states $X(n+1)$ and $X(n)$. Therefore we must represent as state of Y the joint states of both the latter, as shown in the following diagram.



a) Looking at the diagram we can derive the transition matrix $\mathbf{P}$ and find $\mathbf{\Pi}(2)$. Alternatively, starting from state 00 after one step we have $\pi_{00}(1) = 1-p$, $\pi_{10}(1) = p$, and after two steps we have

$$\pi_{00}(2) = (1-p)^2, \qquad \pi_{10}(2) = p(1-p), \qquad \pi_{01}(2) = pq, \qquad \pi_{11}(2) = p(1-q),$$

and remembering the relationship $Y(n+1) = X(n) \oplus X(n+1)$ we finally have for Y

$$\pi_0'(2) = (1-p)^2 + p(1-q), \qquad \pi_1'(2) = p(1-p) + pq.$$

b) Again we must find the asymptotic distribution of chain X_{n+1}, X_n, from the diagram above. By the balance equations we have

$$\pi_{00} = \frac{(1-p)q}{p+q}, \qquad \pi_{10} = \pi_{01} = \frac{pq}{p+q}, \qquad \pi_{11}(2) = \frac{(1-q)p}{p+q},$$

and, hence,

$$\pi_0' = \frac{(1-p)q + (1-q)p}{p+q}, \qquad \pi_1' = \frac{2pq}{p+q}.$$

c) Again looking at the diagram we have

$$\pi_{00}' = \pi_{00,00} + \pi_{11,11} = \frac{(1-p)q}{p+q}(1-p) + \frac{(1-q)p}{p+q}(1-q)$$

$$\pi_{01}' = \pi_{00,10} + \pi_{11,01} = \frac{(1-p)q}{p+q}p + \frac{(1-q)p}{p+q}q$$

$$\pi_{10}' = \pi_{10,11} + \pi_{01,00} = \frac{pq}{p+q}(1-q) + \frac{pq}{p+q}(1-p)$$

$$\pi_{11}' = \pi_{10,01} + \pi_{01,10} = \frac{pq}{p+q}q + \frac{pq}{p+q}p$$

P.1.48 The hour, from 1 to 12, displayed by a clock changes each time by one hour clockwise with probability 1/2, otherwise it does not change.

a) find the asymptotic distribution, if any, of the hour displayed by the clock.

Consider now the case where the clock, in addition to the step clockwise with probability 1/2, is also allowed to step by one hour counterclockwise with probability 1/3. If it starts at 12 at step $n = 0$,

b) find the distribution of the displayed hour at steps $n = 2$ and $n = 3$;

c) verify whether or not the asymptotic distribution, if any, of the hour displayed by the clock is uniform.

Solution a) The balance at node i says $\pi_i = \pi_{i-1}$ for any i. Hence, the asymptotic distribution is $\pi = 1/12$.

b)

$$\mathbf{\Pi}(2) = [1/6 \ 1/4 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1/9 \ 1/9 \ 13/36]$$

c) The balance at node i says

$$\left(\frac{1}{2}+\frac{1}{3}\right)\pi_i = \frac{1}{2}\pi_{i-1} + \frac{1}{3}\pi_{i+1}, \qquad \forall i,$$

which is satisfied if $\pi = 1/12$.

P.1.49 A Markov Chain presents the following one-step transition matrix at all times n.

$$\begin{vmatrix} 0.4 & 0.6 & 0 \\ 0.4 & 0 & 0.6 \\ 0 & 0.4 & 0.6 \end{vmatrix}$$

a) Find the asymptotic (stationary) distribution.

b) Knowing that at time n_0 the distribution of the chain is the stationary one, find the distribution at time $n_0 + 2$.

c) Knowing that at time n_0 the chain is in state 0, find the distribution at time $n_0 + 2$.

Solution

a) $[4/19,\ 6/19,\ 9/19]$

b) If in n_0 is stationary, then it is stationary in any other $n+k$, $k>0$. Therefore the distribution is as in a).

c) $[0.4,\ 0.24,\ 0.36]$.

P.1.50 A binary non-homogeneous Markov Chain $X(n)$ presents the following one-step transition matrices n.

$$\mathbf{P}(2n, 2n+1) = \begin{vmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{vmatrix} \qquad \mathbf{P}(2n+1, 2n+2) = \begin{vmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{vmatrix}$$

for any $n = 0, 1, 2, \ldots$. Starting at time zero with distribution $[1/2,\ 1/2]$ find:

a) the distribution vector at times $n = 2, 3, 4$, and the transition matrices $\mathbf{P}(0,2)$ and $\mathbf{P}(0,3)$;

b) the asymptotic distributions at even time instants, at odd time instants and at any time.

c) the asymptotic joint distribution (all four elements) at two adjacent time instants.

Solution

a) It is easily found that at any time instant, starting from zero, the distribution is uniform, i.e., $[1/2,\ 1/2]$. We also have

$$\mathbf{P}(0,2) = \begin{vmatrix} 0.56 & 0.44 \\ 0.44 & 0.56 \end{vmatrix} \qquad \mathbf{P}(0,3) = \begin{vmatrix} 0.512 & 0.488 \\ 0.488 & 0.512 \end{vmatrix}$$

b) The asymptotic distributions at even time instants is that of the MC whose transition Matrix is $\mathbf{P}(0,2)$ found above. Also, the asymptotic distributions at odd time instants is that of the MC whose transition Matrix is $\mathbf{P}(1,3)$, which turns out equal to $\mathbf{P}(0,2)$. Solving the two balance equations provides a uniform asymptotic distribution in both instants, and then in any instant.

c) We use

$$\pi_{jk} = \pi_j p_{jk}.$$

We must distinguish even and odd instants: π_{jk} is uniform in both cases, p_{jk} not. The matrices are

$$\mathbf{\Pi}(2n, 2n+1) = \begin{vmatrix} 0.3 & 0.2 \\ 0.2 & 0.3 \end{vmatrix} \qquad \mathbf{\Pi}(2n+1, 2n+2) = \begin{vmatrix} 0.4 & 0.1 \\ 0.1 & 0.4 \end{vmatrix}$$

Note that the sum of all elements must be one. The wanted distribution is the average:

$$\mathbf{\Pi}(n, n+1) = \begin{vmatrix} 0.35 & 0.15 \\ 0.15 & 0.35 \end{vmatrix}$$

P.1.51 A Markov Chain presents the following one-step transition matrix at all times n.

$$\begin{vmatrix} 0.4 & 0.6 & 0 \\ 0 & 0.5 & 0.5 \\ 0.3 & 0.3 & 0.4 \end{vmatrix}$$

If the chain starts in $X = 0$ at time $n = 0$, find

a) the transition matrix $\mathbf{P}(0, n)$ and the probability $B(n)$ that the chain is either in state 1 or in state 2 at time $n = 2$;

b) the first order distribution $\mathbf{\Pi}(n)$, and the second order distribution $\mathbf{\Pi}(n, n+1)$ (all terms) at time $n = \infty$;

c) the first order distribution $\mathbf{\Pi}(n)$ at time $n = 4$ knowing that the chain is in state 2 at time $n = 2$; the same, knowing that the chain is in the space subset $(1, 2)$ at time $n = 2$;

d) the first order distribution $\mathbf{\Pi}(n)$ at time $n + 2$ knowing that the chain is in the the space subset $(1, 2)$ at time n with $n = \infty$.

Solution

a)

$$\mathbf{P}(0, 2) = \begin{vmatrix} 0.16 & 0.54 & 0.3 \\ 0.15 & 0.4 & 0.45 \\ 0.24 & 0.45 & 0.31 \end{vmatrix} \qquad \mathbf{\Pi}(2) = [0.16 \;\; 0.54 \;\; 0.3], \qquad B(2) = \pi_1(2) + \pi_2(2) = 0.84.$$

b)

$$\mathbf{\Pi}(n) = [5/27 \;\; 12/27 \;\; 10/27], \qquad \mathbf{\Pi}(n, n+1) = \begin{vmatrix} 2/27 & 3/27 & 0 \\ 0 & 6/27 & 6/27 \\ 3/27 & 3/27 & 4/27 \end{vmatrix}$$

c) Knowing that the chain is in state 2 at time $n = 2$ means $\mathbf{\Pi}(2) = [0 \;\; 0 \;\; 1]$. Therefore

$$[0 \;\; 0 \;\; 1]\mathbf{P}(0, 2) = [0.24 \;\; 0.45 \;\; 0.31].$$

Knowing that the chain is in the space subset $(1,2)$ at time $n=2$ means that the distribution at time $n=2$ is

$$\left[0 \quad \frac{0.54}{0.84} \quad \frac{0.30}{0.84}\right] = [0 \quad 0.6429 \quad 0.3571],$$

i.e., the conditional distribution the Chain has at time $n=2$, starting in state 0 at time 0, conditional to $X(2) \in (1,2)$. The required distribution is

$$[0 \quad 0.6429 \quad 0.3571]\mathbf{P}(0,2) = [0.1821 \quad 0.4179 \quad 0.4000].$$

d) Exactly as question c) where the distribution is equal to $\mathbf{\Pi}(n)$. Knowing that the chain is in subset $(1,2)$, means the conditional distribution is

$$\left[0 \quad \frac{12}{22} \quad \frac{10}{22}\right] = [0 \quad 0.5455 \quad 0.4545],$$

The required distribution is

$$[0 \quad 0.5455 \quad 0.4545]\mathbf{P}(0,2) = [0.1364 \quad 0.4091 \quad 0.4545].$$

P.1.52 A Markov Chain presents the following one-step transition matrix at all times n.

$$\begin{vmatrix} 1/4 & 1/2 & 1/4 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{vmatrix}$$

If the chain starts with distribution $[0.5 \quad 0.5 \quad 0]$ at time $n=0$, find

a) the distribution $\mathbf{\Pi}(2)$ and the asymptotic distribution;

b) the probability that the chain is in state 2 at time $n=3$, and the probability that the chain is in state 1 at time $n=2$ knowing that is in state 2 at time $n=3$.

c) knowing that at time n the chain presents the asymptotic distribution, find the distribution at time $n-1$; then find the distribution at time $n-1$ when at time n the chain is in state 2.

Solution

a)

$$\mathbf{\Pi}(2) = [4/16 \quad 7/16 \quad 5/16], \qquad \mathbf{\Pi} = [3/12 \quad 5/12 \quad 4/12].$$

b)

$$\pi_2(3) = 21/64, \qquad P(X(2)=1|X(3)=2) = 1/3$$

c) If $\mathbf{\Pi}(n)$ is asymptotic, also $\mathbf{\Pi}(n-1)$ is asymptotic. Then we have

$$P(X(n-1)=i|X(n)=2) = \frac{P(X(n)=2)|P(X(n-1)=i)P(X(n-1)=i)}{P(X(n)=2)}$$

Substituting we get

$$[P(X(n-1)=i|X(n)=2)] = [3/16 \quad 5/16 \quad 8/16]$$

P.1.53 A discrete-time random walk is modified in this way. Upon entering a state $i > 0$ the chain remains in the state for three time units, then leaves and with probability p and $1-p$ goes to state $i+1$ and $i-1$ respectively. Leaving state 0, again after three time units, the chain reaches state 1 with probability 1.

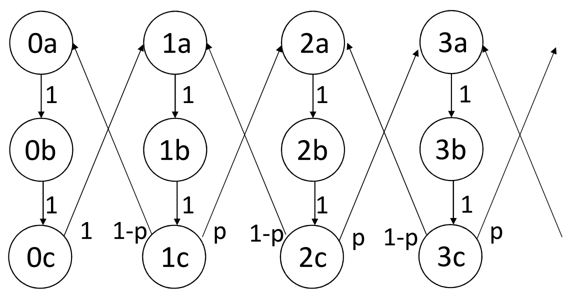
a) Draw the state diagram and find the asymptotic distribution of the chain and the stability conditions on p;

b) the same as the above if the sojourn time is changed to a negative exponential variable with rate μ;

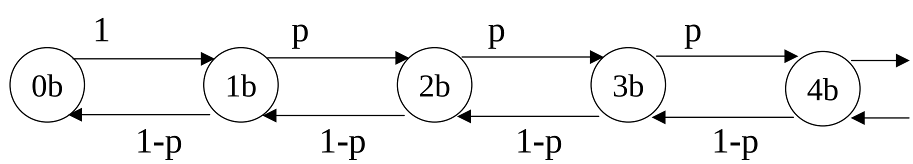
c) the same as point a) if sojourn time in state 0 is changed to one unit time.

Solution

a) The state diagram is as follows



The balance at node b shows $\pi_{ia} = \pi_{ib} = \pi_{ic}$ (The problem is similar to P.1.28 of the class notes, and the solution is the same). In this way the equations reduce to those of the chain in the figure below



whose solution is well known:

$$\pi_{ic} = \pi_{0c} \frac{p^{i-1}}{(1-p)^i}, \qquad i \geq 1,$$

and exists for $1-p > p$, or $p < 0.5$. Then we have

$$\pi_i = \pi_{ia} + \pi_{ib} + \pi_{ic} = \pi_0 \frac{p^{i-1}}{(1-p)^i}, \qquad i \geq 1$$

$$\pi_0 = \frac{1-2p}{2-2p}.$$

b) The asymptotic distribution of a Markov chain only depends on the average of the the sojourn time, and does not change if this average is multiplied by a constant, which is our case, from average 3 to average 1μ.

c) The first diagram above is changed dropping states $0b$ and $0c$, all the remaining being the same. The second figure is still valid if we replace π_{0c} with π_{0a}. We have

$$\pi_{ic} = \pi_{0a}\frac{p^{i-1}}{(1-p)^i}, \qquad i \geq 1,$$

$$\pi_i = \pi_{ia} + \pi_{ib} + \pi_{ic} = 3\pi_{0a}\frac{p^{i-1}}{(1-p)^i} = 3\pi_0\frac{p^{i-1}}{(1-p)^i}, \qquad i \geq 1,$$

$$\pi_0 = \frac{1-2p}{4-2p}.$$

P.1.54 A Markov Chain presents the following one-step transition matrix at all times n.

$$\begin{vmatrix} 0 & 1 & 0 \\ 0.4 & 0 & 0.6 \\ 0 & 1 & 0 \end{vmatrix}$$

If the chain starts in $X = 0$ at time $n = 0$, find

a) the first order distribution $\mathbf{\Pi}(n)$, and the transition matrix $\mathbf{P}(0, n)$ at times $n = 1, 2, 3, 4, \infty$;
b) the second order distribution $\mathbf{\Pi}(n-1, n)$ (all terms) at times $n = 3$ and $n = \infty$;
c) same as question a) starting at time $n = 0$ with the distribution $[0.5 \;\; 0.5 \;\; 0]$;

Solution a) The five transition matrices are as follows

$$\begin{vmatrix} 0 & 1 & 0 \\ 0.4 & 0 & 0.6 \\ 0 & 1 & 0 \end{vmatrix} \quad \begin{vmatrix} 0.4 & 0 & 0.6 \\ 0 & 1 & 0 \\ 0.4 & 0 & 0.6 \end{vmatrix} \quad \begin{vmatrix} 0 & 1 & 0 \\ 0.4 & 0 & 0.6 \\ 0 & 1 & 0 \end{vmatrix} \quad \begin{vmatrix} 0.4 & 0 & 0.6 \\ 0 & 1 & 0 \\ 0.4 & 0 & 0.6 \end{vmatrix} \quad \begin{vmatrix} 0.2 & 0.5 & 0.3 \\ 0.2 & 0.5 & 0.3 \\ 0.2 & 0.5 & 0.3 \end{vmatrix}$$

and the five distributions correspond to the first row of each matrix. Notice that at time $n = \infty$ there is no classical limit and, therefore the result correspond to the generalized concept of limiting distribution.

b) Using matrix positions to represent the join distributions $\mathbf{\Pi}(n-1, n)$ at $n = 1, 2, 3, 4, \infty$, we have (notice that the summation of all elements must be one, while the summation of the rows provides the marginal):

$$\begin{vmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{vmatrix} \quad \begin{vmatrix} 0 & 0 & 0 \\ 0.4 & 0 & 0.6 \\ 0 & 0 & 0 \end{vmatrix} \quad \begin{vmatrix} 0 & 0.4 & 0 \\ 0 & 0 & 0 \\ 0 & 0.6 & 0 \end{vmatrix} \quad \begin{vmatrix} 0 & 0 & 0 \\ 0.4 & 0 & 0.6 \\ 0 & 0 & 0 \end{vmatrix} \quad \begin{vmatrix} 0 & 0.2 & 0 \\ 0.2 & 0 & 0.3 \\ 0 & 0.3 & 0 \end{vmatrix}$$

c) The five matrices $\mathbf{P}(0, n)$ are, by definition, the same as in a), while vectors $\mathbf{\Pi}(n)$ are given by the average of the first two rows of the $\mathbf{\Pi}(n)$ and are all equal to

$$[0.2 \;\; 0.5 \;\; 0.3]$$

Chapter 2

More on arrivals and discrete-state processes

2.1 Multi-dimensional Poisson Events

The definition of Poisson Events on the time axis (Poisson Arrivals) can easily be extended to two or more dimensions. In this context, the time coordinate t is replaced by coordinate vector $\mathbf{x} = \{x_1, \dots, x_n\}$.

Let start with Poisson Events on the plane. Denoted by ΔS the infinitesimal measure of an infinitesimal surface, the three Poisson axioms are re-formulated in the following way:

first Axiom (2.1)

$$P(N(\Delta S) = 1) = \sigma \Delta S + \omega(\Delta S);$$

second Axiom (2.2)

$$P(N(\Delta S) > 1) = \omega(\Delta S);$$

third Axiom (2.3)
The number of events that occur in disjpiont areas arev statistically independent RVs.

In the same way as we did on the line we can show that the distribution of $N(S)$, the RV number of events occurring in an area of measure S, is

$$P(N(S) = k) = \frac{(\sigma S)^k}{k!} \, e^{-\sigma S}, \qquad k \geq 0 \tag{2.4}$$

independently of the location of S in the plane, where σS is the average number of events in S and σ is the density.

The Poisson model is used to represent phenomena such as the distribution of the seeds on a surface, the distribution of stars in space, etc..

Example (2.5)
Find the ddp of RV Z, the distance from a given point O of the Poisson event closest to O.

We have

$$f(z)\Delta z = P(N(S) = 0)\ P(N(\Delta S) = 1)$$

where S is the measure of the surface of the circle of radius z, and Δ S is the increase of this surface when the radius is increased by Δz. You get:

$$f_Z(z) = 2\pi\sigma z\ e^{-\sigma\pi z^2}, \qquad z \geq 0, \tag{2.6}$$

a pdf known as Rayleigh's pdf.

In three dimensions, we have similarly:

$$f_Z(z) = 4\pi\rho z^2\ e^{-4/3\ \rho\pi z^3} \qquad z \geq 0 \tag{2.7}$$

where ρ represent the density of the events. The above is known as Maxwell's pdf.

Example (2.8)
The flying bombs that fell in the south of London in the World War II adapted very well, as we discovered later, to the Poisson model with density $\sigma = 3.729$ bombs/km^2. Find the probability that $0, 1$ *or more bombs could hit an area of 1000 m^2 and 1 km^2.*

We use

$$P(N(s_a) = k) = \frac{(\sigma s_a)^k}{k!}\ e^{-\sigma s_a}$$

that yields

	$k = 0$	$k = 1$	$k \geq 2$
$s_a = 10^{-3}$ km^2	0,9962 ...	0,0037 ...	0,0000 ...
$s_a = 1$ km^2	0,0240 ...	0,0895 ...	0,8864 ...

Example (2.9)
Given Poisson points in the plane xy with density σ, we denote by A_n those points that fall at a distance from x less than or equal to d, and with A'_n their projections on the x axis. Find the probability distribution of RV $N'(\ell)$, the number of a'_n points that lie in a segment of length ℓ of the x axis (Figure 2.1).

$P(N'(\ell) = k)$ is the probability that k Poisson events in the plane lie within a rectangle R of sides ℓ and $2d$:

$$P(N'(\ell) = k) = P(N(R) = k) = \frac{(2\sigma d\ell)^k}{k!}\ e^{-2\sigma d\ell}$$

Points A'_n represent Poisson events on the line whose density is $\lambda = 2\sigma d$.

With reference to the previous example, points A'_n may represent interruptions in telephone lines or other ducts due to the bombs that fall within a distance d. With previous numerical values and setting $d = 50$ m you get $\lambda = 0.186\ldots$ interruptions/km.

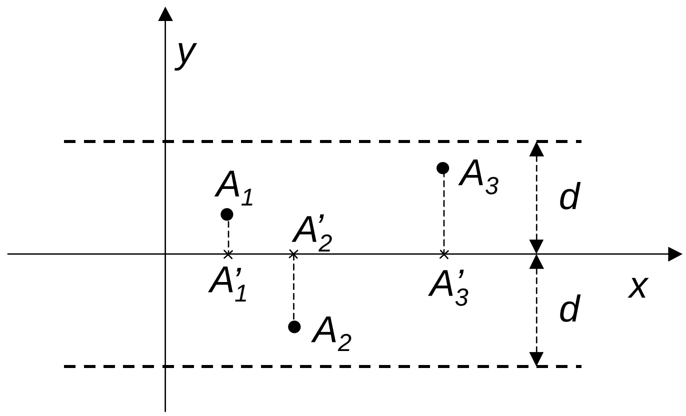


Figure 2.1:

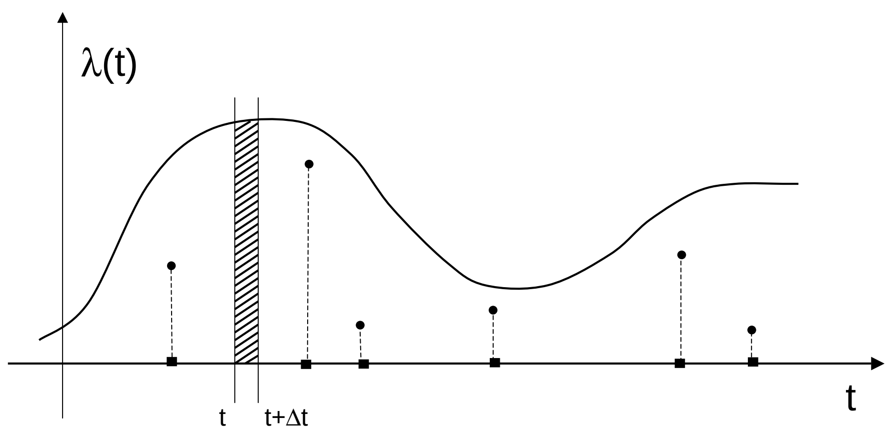


Figure 2.2:

2.1.1 Non-homogeneous Poisson Arrivals

If the first axiom of Poisson arrivals is modified so as to allow the time dependence of the parameter $\lambda = \lambda(t)$, then we get non-homogeneous Poisson arrivals.

These events can be made to derive in a simple way by Poisson Events in the plane. Let, in fact, consider the plane $\{\lambda(t), t\}$, in which thare are Poisson Events in the plane with density $\sigma = 1$ (Figure 2.2). The points represented by the projection on t of the two-dimensional Events that lie between curve $\lambda(t)$ and t-axis, are the sought events. The probability that we have a non-homogeneous event in $[t, t + \Delta t]$, equals the probability that a two-dimensional event lies in the elementary strip shown in the figure. By the first axiom applied to two-dimensional Events we have, therefore

$$P(N(t, t + \Delta t) = 1) = \sigma\lambda(t)\Delta t = \lambda(t)\Delta t \tag{2.10}$$

Denoted

$$M(t,\tau) = \int_t^{t+\tau} \lambda(t)dt \tag{2.11}$$

the same argument as above shows that

$$P(N(t,t+\tau)=k) = \frac{M(t,\tau)^k}{k!}\ e^{-M(t,\tau)}, \qquad k \geq 0, \tag{2.12}$$

which shows that the number of Events in $[t\,; t+\tau]$ still presents a Poisson distribution with average $M(t,\tau)$.

The pdf of the RV $V(t)$, waiting time to the next event, can be derived with the usual argument:

$$f_{V(t)}(x,t) = \lambda(t+x)\ e^{-M(t,x)}, \qquad x \geq 0. \tag{2.13}$$

The dependance on t shows that $V(t)$ is no longer stationary, although it is still markovian.

Example (2.14)
A non-homogeneous Poisson Arrival process has an arrival rate $\lambda(t)$ with a periodic behavior that alternates at every period T, two constant rates, namely λ_a (phase a) and λ_b (step b).

Such a process presents an asymptotic distribution that is the one we observe in a Random Inspection Point (RIP) t_1. As this point may equally likely lie in one of two periods, we have:

$$P(N(t_1,t_1+\Delta t)=1) = \lambda\Delta t = \frac{\lambda_a+\lambda_b}{2}\Delta t.$$

Also time $t_2 = t_1 + \Delta t$ represents a RIP, and the relation above still holds with the substitution of t_1 with t_2. In the limit $\Delta t \to 0$, t_2 belongs with probability one to the same phase as t_1, and therefore we have

$$P(N(t_1,t_1+\Delta t_1)=1; N(t_2,t_2+\Delta t_2)=1) = \frac{1}{2}\lambda_a^2\Delta t_1\Delta t_2 + \frac{1}{2}\lambda_b^2\Delta t_1\Delta t_2 = \frac{\lambda_a^2+\lambda_b^2}{2}\Delta t_1\Delta t_2,$$

which is different from

$$P(N(t_1,t_1+\Delta t_1)=1)P(N(t_2,t_2+\Delta t_2)=1) = (\frac{\lambda_a+\lambda_b}{2})^2\Delta t_1\Delta t_2.$$

As a consequence we have the apparent paradox that, while the considered process is Poisson, and therefore, a Purely Random process, even though not stationary, asymptotically we observe stationary conditions, but not a Purely Random process! ♣

2.1.2 Compound Poisson Arrivals

These Events can be derived by Poisson Arrivals by relaxing the second axiom and allowing more than one arrivals in Δt. In practice, the first two axioms are replaced by the axioms that govern multiple arrivals of order i:

Axiom (2.15)

$$P(N(t,t+\Delta t)=i)=\lambda_i\Delta t+\omega(\Delta t), \qquad i>0.$$

We can also view the multiple arrival as a single event (compuond event) whose rate is given by

$$\lim_{\Delta t\to 0}\frac{P(N(t,t+\Delta t)>0)}{\Delta t}=\lambda$$

with

$$\lambda=\sum_{i=1}^{\infty}\lambda_i,$$

whereas the average rate of the arrivals is

$$\sum_{i=1}^{\infty} i\lambda_i.$$

The probability that a single compound event is composed by $N_c=i$ arrivals can be written as

$$P(N_c=i)\equiv g_i=\frac{\lambda_i}{\lambda} \qquad i>0,$$

a conditional probability.

Denoted by N_i the number of compound events, in τ, each composed of i arrivals, we have

$$P(N_1=k_1;N_2=k_2;\ldots;N_i=k_i;\ldots)=P(N_1=k_1)P(N_2=k_2);\ldots;P(N_i=k_i);\ldots=$$

$$=\frac{(\lambda_1\tau)^{k_1}}{k_1!}\,e^{-\lambda_1\tau}\frac{(\lambda_2\tau)^{k_2}}{k_2!}\,e^{-\lambda_2\tau}\ldots\frac{(\lambda_i\tau)^{k_i}}{k_i!}\,e^{-\lambda_i\tau}\ldots=$$

$$=e^{-\lambda\tau}\frac{(\lambda_1\tau)^{k_1}}{k_1!}\frac{(\lambda_2\tau)^{k_2}}{k_2!}\ldots\frac{(\lambda_i\tau)^{k_i}}{k_i!}\ldots \tag{2.16}$$

The distribution of the number N of arrivals in τ is rather complex to determine, since the occurrence of k arrivals can correspond to a variety of compound events.

Example (2.17)
Find the probability that we have three arrivals in τ.

We have

$$P(N(\tau)=3)=P(N_1=3;N_i=0,i>1)+P(N_1=1;N_2=1;N_i=0,i>2)+$$

$$+P(N_3=1;N_i=0,i\neq 3)=e^{-\lambda\tau}\left(\frac{(\lambda_1\tau)^3}{3!}+\lambda_1\tau\lambda_2\tau+\lambda_3\tau\right)\clubsuit$$

By the definition the distance between two consecutive type i events has a negative exponential pdf with rate λ_i, while the distance between two consecutive non-coincident events of any type has a negative exponential pdf with rate $\sum_i\lambda_i$.

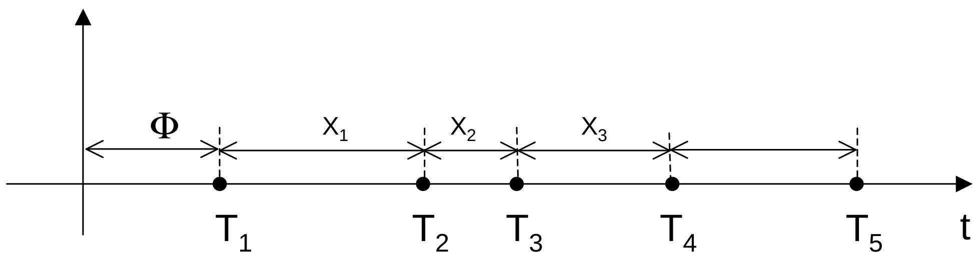


Figure 2.3:

2.2 Renewal Events

Regenerative Arrivals in which every arrival is a regenerative instant are called Renewal Events. By definition, the distances $X_1, X_2, \dots X_n, \dots$ between consecutive events are independent and with the same pdf $f_X(x)$. Therefore, we will refer to X as the *renewal variable*. The location of Renewal Events on the arrival axis is provided by RV Φ, coordinate of the first arrival, and called initial delay. In particular, both Bernoulli and Poisson arrivals belong to the renewal family, for which Φ coincides with X (1.18).

As the Renewal Variable X coincides with the cycle length of the regenerative process, Renewal Events are called *positive recurrent*, *null recurrent* and *non-recurrent*, as in regenerative processes.

In the following we refer to positive-recurrent events, and to the continuous-time case, from which the properties in the discrete-time case are readily derived.

The arrival rate

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(N(t, t + \Delta t) = 1)}{\Delta t}$$

is called the *renewal rate*. Unfortunately, it is very difficult to find the renewal rate also with non-delayed events. In fact this event is the union of the following infinite events

$$\{t \le X_1 \le t + \Delta t\} + \{t \le X_1 + X_2 \le t + \Delta t\} + \{t \le X_1 + X_2 + X_3 \le t + \Delta t\} + \dots$$

Nevertheless, we will derive interesting asymptotical properties for positive recurrent renewals.

2.2.1 The conditional renewal rate

Here we are interested in deriving the renewal rate at time t, conditional to the knowledge of the whole past history, i.e.,

$$\beta(t, |\text{past history}) = \lim_{\Delta t \to 0} \frac{P(N(t, t + \Delta t) = 1 | \text{past history})}{\Delta t}$$

The renewal paradigm implies that, once we know that the event immediately prior to t happened τ units backward, i.e., at time $t - \tau$, then what happened in the past up to time $t - \tau$ does not matter. This changes the expression of the renewal probability as

$$P(N(t, t + \Delta t) = k|\text{past history}) = P(N(t, t + \Delta t) = k|D = \tau)$$

where RV D represents the time since the last renewal before t. This makes the *conditional renewal rate* independent of t and dependent only on τ. This is the more evident when we express event in terms of the Renewal Variable X, i.e.

$$\{P(N(t, t + \Delta t) = 1|D = \tau)\} \equiv \{\tau < X \leq \tau + \Delta t|X > \tau\}$$

Therefore, the definition of the conditional renewal rate becomes

$$\beta(\tau) = \lim_{\Delta t \to 0} \frac{P(\tau < X \leq \tau + \Delta t|X > \tau)}{\Delta t} = \lim_{\Delta t \to 0} \frac{P(\tau < X \leq \tau + \Delta t; X > \tau)}{P(X > \tau)\Delta t}. \tag{2.18}$$

Being in the limit

$$P(\tau < X \leq \tau + \Delta t) = f_X(\tau)\Delta t$$

the stationary conditional renewal rate appears to be related to the pdf of the renewal RV X as follows:

$$\beta(\tau) = \frac{f_X(\tau)}{1 - F_X(\tau)} \tag{2.19}$$

Note that the rate does not present the characteristics of a pdf; in fact the variation of t also changes condition. Moreover, the area of $\beta(\tau)$ diverges and this is a necessary condition for a function to represent a rate.

Example (2.20)
find the conditional renewal rate when X is a negative exponential RV (Poisson Arrivals).

$$\beta(\tau) = \frac{f_X(\tau)}{1 - F_X(\tau)} = \frac{\lambda e^{-\lambda\tau}}{e^{-\lambda\tau}} = \lambda$$

The conditional rate equals the unconditional one. This was expected since Poisson Arrivals are purely random, and the occurrence of an arrival does not depend on past history whatsoever. For this reason the exponential RV is said *memoryless*. This highlights the link between the exponential and the pure randomness.♣

Example (2.21)
find the conditional renewal rate when X is a uniform RV in $[0 : T]$.

$$\beta(\tau) = \frac{f_X(\tau)}{1 - F_X(\tau)} = \frac{1/T}{(T - \tau)/T} = \frac{1}{(T - \tau)}, \qquad 0 \leq \tau \leq T$$

Note that with the increase of τ the rate increases and diverges in $\tau = T$, because the renewal must necessarily happen within T (Figure2.4).♣

Now we derive the inverse relation noting that

$$\int_0^t \beta(\tau)d\tau = -\ln(1 - F_X(t))$$

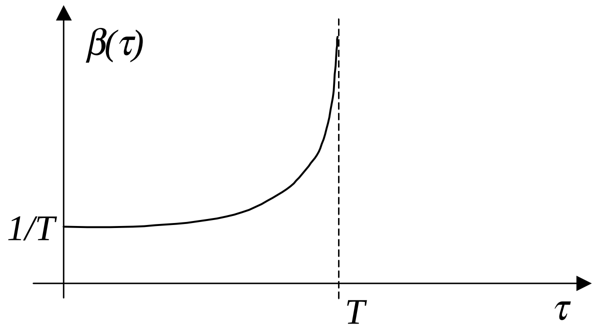


Figure 2.4:

and, therefore,

$$F_X(t) = 1 - e^{-\int_0^t \beta(\tau)d\tau}$$

$$f_X(t) = \beta(t)e^{-\int_0^t \beta(\tau)d\tau} \tag{2.22}$$

Example (2.23)
It left to the reader to show that the pdf corresponding to a rate proportional to the time ($\beta(\tau) = \alpha\tau$), provides a Rayleigh pdf.

If X is an integer RV, *memoryless* implies the geometric distribution.

2.2.2 Renewal laws

In this section we derive some important properties regarding positive recurrent renewals. Because for positive recurrent regeneration processes an asymptotic distribution always exists (Theorem 1.185), we can define the asymptotic, or stationary, renewal rate as

$$\lim_{t\to\infty} \lambda(t) = \lambda$$

Theorem: *First Renewal Theorem* (2.24)
For positive recurrent renewal processes, the asymptotic renewal rate λ and the average renewal variable m_X, are related by

$$\lambda = 1/m_X. \tag{2.25}$$

The thesis comes from Theorem 1.185. Indeed, process $N(t, t+\Delta t)$ everywhere zero and equal to one only within intervals Δt that contains the renewal event, and we asymptotically have, using Theorem 1.185, $E[Y_1] = \Delta t$ and $E[C] = m_X$, that provides

$$P(N(t, t+\Delta t) = 1) = \frac{\Delta t}{m_X}.$$

However, we also have

$$P(N(t, t + \Delta t) = 1) = \lambda \Delta t \quad \clubsuit$$

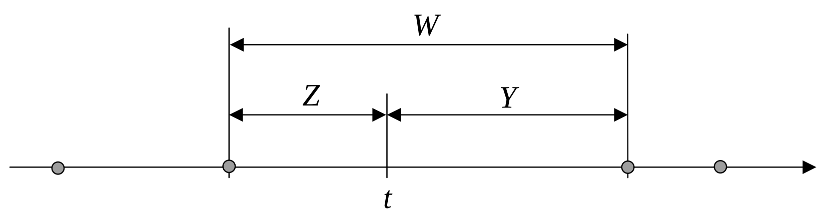


Figure 2.5:

Figure 2.5, shows renewal events and RVs linked to them, that lie around a chosen instant t. The RV W is the distance between the two events that lie across t. riding on t. It is obviously linked to V.C. X, the distance between the renewal events, but we have changed the symbol because it now is not the distance between a randomly selected event and its successor, but, rather, is the distance between two consecutive events that lie across t. The experiment that leads to the two RVs are different, and, therefore, the RVs are different.

Unfortunately, it is almost impossible to derive the pdf of W in t, except when t is an asymptotic point in the broad sense. We will derive it using the RIP method.

Letting m_X, σ_X^2 and m_{2X} respectively the average, the variance, and the second order moment of X, we have

Theorem: *The Renewal Paradox I* (2.26)
The length W of the renewal interval that lies across an asymptotic time instant t has pdf:

$$f_W(x) = \frac{x f_X(x)}{m_X} \tag{2.27}$$

$$E[W] = m_X + \frac{\sigma_X^2}{m_X} = \frac{m_{2X}}{m_X}$$

We see that the pdf of W is broader than the pdf of X, as we can see also from its mean. This is because the RIP lies more easily in larger renewal intervals.

Proof
Take at first a very large interval T, expected to contain many renewal intervals. Take one of such intervals and denote by E the event: $\{$*the RIP lies in the selected interval*$\}$. By using Bayes theorem we have

$$f_W(x) = f_X(x|E) = \frac{P(E|X = x) f_X(x)}{\int P(E|X = x) f_X(x) dx}$$

By observing that $P(E|X = x) = x/T$, we get

$$f_W(x) = \frac{x/T \ f_X(x)}{\int x/T \ f_X(x) dx}$$

The above relation does not depend on T and proves the thesis. ♣

Theorem: *The Renewal Paradox II* (2.28)
The waiting time Y to the next renewal event in an asymptotic time instant has pdf:

$$f_Y(x) = \frac{1 - F_X(x)}{m_X} \tag{2.29}$$

$$E[Y] = \frac{m_X}{2} + \frac{\sigma_X^2}{2m_X}$$

Proof
Since the RIP is uniformly taken, if we know $W = w$ the conditional pdf of Y is uniform in $[0; w]$:

$$f_Y(x|w) = \frac{1}{w} \qquad 0 \leq x \leq w$$

By the Total Probability Theorem applied to (2.27), we have:

$$f_Y(x) = \int_x^\infty \frac{w f_X(w)}{m_X} \frac{1}{w}\, dw = \frac{1 - F_X(x)}{m_X} \qquad ♣$$

By symmetry, referring to Figure 2.5 we have, for RV Z

Corollary (2.30)
$f_Z(z) = f_Y(z)$.

Example (2.31)
Let apply (2.27)(2.29) to Poisson Arrivals, where X is negative exponential.

We have

$$f_W(x) = \frac{x f_X(x)}{m_X} = \frac{x\lambda e^{-\lambda x}}{1/\lambda} = \lambda^2 x e^{-\lambda x}, \qquad x \geq 0$$

La V.C. W è dunque una Erlang-2. Si ha poi

$$f_Y(x) = \frac{1 - F_X(x)}{m_X} = \frac{e^{-\lambda x}}{1/\lambda} = \lambda e^{-\lambda x}, \qquad x \geq 0$$

Y still presents the same pdf as X. This is because of the memoryless property of the negative exponential pdf, we have already seen. Variable W is Erlang-2, in agreement to the former comment.♣

Example (2.32)
Let apply (2.27)(2.29) to periodic arrivals with period T.

We have

$$f_W(x) = \frac{x f_X(x)}{m_X} = \frac{x\delta(x - T)}{T} = \delta(x - T)$$

RV W is still equal to the constant T (Obviously). Then,

$$f_Y(x) = \frac{1 - F_X(x)}{m_X} = \frac{1 - u(x - T)}{T} = \frac{1}{T} \qquad 0 \le x \le T$$

and Y is uniform within $[0; T]$.♣

Example (2.33)
As the previous examples when X is Erlang-2 with pdf (1.35).

The CDF of Erlang-k is easily attained remembering its relationship with Poisson Arrivals at rate λ:

$$F_X(x) = P(X \le x) = 1 - P(X > x) = 1 - P(A(x) \le k - 1)$$

being $A(x)$ the number of Poisson Arrivals in x. Then,

$$F_X(x) = \begin{cases} 1 - \sum_{i=0}^{k-1} \frac{(\lambda x)^i}{i!}\, e^{-\lambda x} & (x \ge 0) \\ 0 & (x < 0) \end{cases} \tag{2.34}$$

Equation (2.29) provides

$$f_Y(y) = \frac{1}{2}\lambda e^{-\lambda y} + \frac{1}{2}\lambda^2 y e^{-\lambda y}$$

Note that the result represents the arithmetic average between the exponential and the Erlang-2, pdfs', and this is explained very well if we remember that the Erlang-2 is the distance between arrival i and arrival $i + 2$. With probability 1/2 the RIP lies between arrivals $i + 1$ and $i + 2$ (Y exponential), and with probability 1/2 the RIP lies between i and $i + 1$ (Y Erlang-2).

As for W, we have:

$$f_W(x) = \frac{x f_X(x)}{m_X} = \frac{x\lambda^2 x e^{-\lambda x}}{2/\lambda} = \frac{(\lambda x)^2}{2}\lambda e^{-\lambda x}, \qquad x \ge 0,$$

i.e., W is Erlang-3 (the RIP behave exactly as a Poisson arrival). ♣

From Theorem 2.28 we can also re-derive Theorem 2.24. In fact, if t is the RIP, the preceding theorem assures that

$$\lambda \Delta t = P(N(t, t + \Delta t) = 1) = f_Y(0)\Delta t = \frac{1 - F_X(0)}{m_X}\Delta t = \frac{1}{m_X}\Delta t.$$

Also, Theorem 2.28 suggests how to take the initial delay Φ in order to build Stationary Renewal Events. In fact, the (2.29) must hold in any stationary point, even when t is finite. So we have

Corollary (2.35)
A sufficient condition for Renewal Events to be stationary is that the pdf of the initial delay Φ equals (2.29).

2.2.3 Case study

In general it is not easy the derive even the first order statistic $\lambda(t)$. An exception is the case where the Renewal Interval has pdf Erlang-2 with no initial delay. This case is made simple by its connection with Poisson arrivals; in fact, sech renewal events are derived by Poisson ones dropping the Poisson arrival alternately, so that only arrivals with an even order number survive.

Denoted by $N_s(t, t+\Delta t)$ the number of surviving arrivals in $[t; t+\Delta t]$ we have

$$\begin{aligned} P(N_s(t,t+\Delta t)=1) &= P(N(t,t+\Delta t)=1)P(\text{an odd number of arrivals in } [0;t]) = \\ &= P(N(t,t+\Delta t)=1)\sum_{k=0}^{\infty} P(N(0,t)=2k+1) \\ &= \lambda\Delta t \sum_{k=0}^{\infty} \frac{(\lambda t)^{2k+1}}{(2k+1)!} e^{-\lambda t} \end{aligned} \tag{2.36}$$

The summation $_o$ of the odd terms above can be evaluated together with S_e the summation of even terms in the following way:

$$\begin{cases} S_e + S_o = 1, \\ \\ S_e - S_o = \displaystyle\sum_{k=0}^{\infty} (-1)^k \frac{(\lambda t)^k}{k!}\ e^{-\lambda t} = e^{-2\lambda t}, \end{cases}$$

$$\begin{cases} S_e = \dfrac{1+\ e^{-2\lambda t}}{2}, \\ \\ S_o = \dfrac{1-\ e^{-2\lambda t}}{2}. \end{cases}$$

Therefore we have

$$P(N_s(t,t+\Delta t)=1) = \frac{1-\ e^{-2\lambda t}}{2}\lambda\Delta t. \tag{2.37}$$

We see that the process is not stationary, i.e., the first Poisson axiom is not valid. It also happen that the third Poisson axiom is also not valid. In fact, we have, for example,

$$P(N_s(t,t+\Delta t)=1; N_s(t+\Delta t, t+2\Delta t)=1) = 0,$$

since, by hypothesis, the two events can not occur for $\Delta t \to 0$, since referring to the Poisson flow, one of the two is surely dropped. If there would be idependence, then we would have

$$P(N_s(t,t+\Delta t)=1; N_s(t+\Delta t, t+2\Delta t)=1) = \lambda_1\Delta t\lambda_2\Delta t = \lambda_1\lambda_2(\Delta t)^2$$

2.2.4 Merge of Renewal Events

In general, if you merge on different Renewal Events, which below assume stationary and independent, on the same time axis you do not get Renewal Events. We can easily verifY this when we merge two such processes. This results has one notably exception, i.e., when the merged events are Poisson, as we have already seen dealing with Poisson processes.

As it happens for the Poisson events, the instantaneous rate λ of the composition of events equals the sum of the rates of the cpmponent Events:

$$\lambda = \sum_{i=1}^{n} \lambda_i \tag{2.38}$$

Let assume now that we increase the number n of component events, and reduce the rate λ_i of each event in a manner inversely proportional to n, so that the total rate $\sum \lambda_i$ remains unchanged. This means that the rate λ_i of each component tends to zero so such that:

$$\lim_{n\to\infty} \frac{n\lambda_i}{\lambda} \ = \ 1 \tag{2.39}$$

We have:

Theorem: (2.40)
Events resulting from the composition of an infinite number of Renewal events asymptotically constitute Poisson events.

Proof
You only need to show that events that happens in two or more small and not overlapping intervals $\Delta t_1,\ \Delta t_2, \ldots$, are statistically independent (third axiom of Poisson, since the first two are already ‘a valid).

This is of immediate intuition when we consider that, when n is finite, the dependence is due to the memory between arrivals (general interarrival RV are not memoryless) that exists in each of the component processes. However, when the average frequency of the single components tends to zero, no component contributes with more than one event, and this cancel the dependence.

The formal proof along this way is complicated by the fact we need at least the joint pdfs' of the second order.♣

To end this section about Regenerative processes we must make known that almost all the theorems we have proven here make no use of the independence of cycles. This means that such theorems also apply to a broader class of processes (see later on the Ergodic Processes), even though this class can not be easily defined.

2.3 Problems for solution (on renewal events)

P.2.1 Find the conditional renewal rate when the renewal interval has an Erlang-2 pdf with parameter λ. Discuss such rate when we take the limits $\tau \to 0$ and $\tau \to \infty$.

P.2.2 Consider Renewal Events where the renewal variable has pdf Erlang-2 with parameter λ. In every renewal interval we insert a random event in an uniform way. Is the new point process a Renewal Process? If so, find the pdf of the new renewal RV.

P.2.3 Repeat the problem above when points are added at exactly the center of the old renewal interval X.

P.2.4 (2.21) At a bus stop busses arrive according to a Poisson process of rate μ. Passengers arrive at the bus stop according to a Poisson process with rate λ. A bus arrives at the stop and finds one passenger.

a) find the pdf of the time elapsed since the last bus arrived.

b) as in a) when busses arrive according to a renewal process with interarrival pdf $f_X(x)$ and $\lambda \to 0$.

P.2.5 (2.26) Arrivals occur with distances that are independent negative exponential RV, but with rates that alternate, λ and μ. Find

a) the rate of arrival (be careful, the result is not the ...)

b) chosen an event, the pdf of the distance to the next arrival;

c) taken a Random Inspection Point, the pdf of the waiting time to the next arrival

P.2.6 (2.23) Arrivals occur at a distance that alternates between RVs X and Y, all samples being independent.

a) Find the asymptotic rate of arrival;

b) taken a Random Inspection Point, find the probability that it lies within an interval of type X;

c) The pdf of the waiting time to the next arrival.

P.2.7 (2.40) Arrivals Occur in a periodic way, alternating constant intervals of value of T and $2T$. Taken a Random Inspection Point, find

a) the probability that the RIP lies in an interval of length $2T$;

b) the pdf of the waiting time to the next arrival;

c) the probability an arrival occurs in the next Δt;

P.2.8 (2.41) Arrivals occur according to Renewal Events where the interarrival period has pdf Erlang-k, $k > 1$ with parameter λ. Assuming that the first arrival is in the origin (t=0), find $P(A(t, t + \Delta t) = 1)$, disregarding infinitesimals $(\Delta t)^k$ of order $k > 1$, in the two cases

a) $t = 0$;

b) $t \to \infty$.

and compare with the case $k = 1$.

c) Find the probability that the first arrival after the origin occurs in $[t, t + \Delta t]$, in the two cases $k = 1, 2$.

P.2.9 Given Renewal Arrivals of rate λ, find the probability $P(A(t, t + \Delta t) = 1, A(t + \Delta t, t + 2\Delta t) = 1)$ where t is a Random inspection point and the pdf $f_X(x)$ of the interarrival time is

a) Negative exponential

b) Erlang-2.

b) Uniform, starting from 0.

Note that the parameters of pdfs' above can be derived by the rate of the arrivals λ

P.2.10 The same as the previous Problem where now the probability to be found is $P(A(t, t + \Delta t) = 0, A(t + \Delta t, t + 2\Delta t) = 1)$

P.2.11 (3.22) Poisson arrivals are deleted according to the following procedure: arrival i is suppressed with probability p if arrival $i - 1$ has been suppressed, otherwise it is suppressed with probability q.

a) Find the asymptotic probability ($i \to \infty$) of suppression of an arrival (Hint: Consider the discrete-time binary process S_i *1 non-suppression, 0 suppression.* This process carries memory, ...);

b) Taken a survived arrival i, find the probability P_k that point $i + k$ is the first survived after i (the answer does not depend on the answer to point a); take at first $k = 1$ and then...);

c) find the average distance between survived arrivals.

2.1 Problems' solutions- Chapter 2

P.2.1 Find the conditional renewal rate when the renewal interval has an Erlang-2 pdf with parameter λ. Discuss such rate when we take the limits $\tau \to 0$ and $\tau \to \infty$.

Solution.
By the definition we get

$$\beta(\tau) = \frac{\lambda^2 \tau}{1 + \lambda \tau}$$

Finally we have:

$$\lim_{\tau \to 0} \beta(\tau) = 0$$

and

$$\lim_{\tau \to \infty} \beta(\tau) = \lambda$$

Why?

P.2.2 Consider Renewal Events where the renewal variable has pdf Erlang-2 with parameter λ. In every renewal interval we insert a random event in an uniform way. Is the new point process a Renewal Process? If so, find the pdf of the new renewal RV.

Solution.
The old renewal interval X is split into two intervals Z and Y, which, by symmetry, have the same pdf. We have:

$$f_Z(z) = \int f_z(z/X = x) f_X(x) dx = \int_z^{\infty} \frac{1}{x} \lambda^2 x e^{-\lambda x} dx = \lambda e^{-\lambda z}$$

We see that the new events are of the Poisson type.

P.2.3 Repeat the problem above when points are added at exactly the center of the old renewal interval X.

Solution- The new interval is exactly $Z = X/2$ where X is the old one. So, Z is again an Erlang-2 RV where the average is halved.

$$f_X(x) = 4\lambda^2 x e^{-2\lambda x} \qquad x \geq 0$$

P.2.4 (2.21) At a bus stop busses arrive according to a Poisson process of rate μ. Passengers arrive at the bus stop according to a Poisson process with rate λ. A bus arrives at the stop and finds one passenger.

a) find the pdf of the time elapsed since the last bus arrived.

b) as in a) when busses arrive according to a renewal process with interarrival pdf $f_X(x)$ and $\lambda \to 0$.

Solution.
a) Using Bayes' Theorem:

$$f_X(x/N(x)=1) = \frac{P(N(X)=1/X=x)f_X(x)}{\int P(N(X)=1/X=x)f_X(x)dx)} = \frac{\lambda x e^{-\lambda x}\mu e^{-\mu x}}{\int_0^\infty \lambda x e^{-\lambda x}\mu e^{-\mu x}dx} =$$

$$= (\lambda+\mu)^2 x e^{-(\lambda+\mu)x}$$

. Erlang-2 with parameter $\lambda + \mu$.

b) we have

$$f_X(x/N(x)=1) = \frac{\lambda x e^{-\lambda x} f_X(x)}{\int_0^\infty \lambda x e^{-\lambda x} f_X(x)dx}$$

Se $\lambda \to 0$ si ha

$$f_X(x/N(x)=1) = \frac{x f_X(x)}{m_x}$$

This is the same result as the Renewal Paradox. This shows again that a Poisson arrival with $\lambda \to 0$ behaves exactly like a Random Inspection Point

P.2.5 (2.26) Arrivals occur with distances that are independent negative exponential RV, but with rates that alternate, λ and μ. Find

a) the rate of arrival (be careful, the result is not the ...)
b) chosen an event, the pdf of the distance to the next arrival;
c) taken a Random Inspection Point, the pdf of the waiting time to the next arrival

Solution
a) Arrivals are not Poisson nor renewals; however, the arrival process is regenerative, where the cycle length is the summation of two consecutive interarrival times. Therefore we can apply the limiting theorem for these processes, as we did for renewals, and we get

$$P(N(t, t+\Delta t) = 1 = \frac{2\Delta t}{m_X + m_Y}$$

and the rate is

$$\eta = \frac{2}{m_X + m_Y} = \frac{2\lambda\mu}{\lambda+\mu}$$

b) An arrival is of specific kind with probability 1/2. Therefore, by the Total Probability Theorem we have:

$$f_X(x) = \frac{1}{2}\lambda e^{-\lambda x} + \frac{1}{2}\mu e^{-\mu x}$$

c) The probability that the RIP lies within an interval of average $1/\lambda$ and $1/\mu$ is respectively

$$\frac{\mu}{\lambda+\mu}, \qquad \frac{\lambda}{\lambda+\mu}.$$

Thus we have:

$$f_Y(y) = \frac{\mu}{\lambda+\mu}\lambda e^{-\lambda x} + \frac{\lambda}{\lambda+\mu}\mu e^{-\mu x}$$

P.2.6 (2.23) Arrivals occur at a distance that alternates between RVs X and Y, all samples being independent.

a) Find the asymptotic rate of arrival;
b) taken a Random Inspection Point, find the probability that it lies within an interval of type X;
c) The pdf of the waiting time to the next arrival.

Solution.
Arrivals do not strictly belong to Renewal Events, however the process is still regenerative, being the cycle equal to $X+Y$. Therefore we can apply the limiting theorem for these processes, as we did for renewals, and we get

$$P(N(t, t+\Delta t) = 1 = \frac{2\Delta t}{m_X + m_Y}$$

and the rate is

$$\lambda = \frac{2}{m_X + m_Y}$$

the reverse of the average interarrival period,

$$m = \frac{1}{2}m_X + \frac{1}{2}m_Y,$$

exactly as it happens for renewals.

b) Again, this is regenerative, and with the same thorem as above we see that

$$P(in\ X) = \frac{m_X}{m_X+m_Y}, \qquad P(in\ Y) = \frac{m_Y}{m_X+m_Y}.$$

c) Denoted by $f_Z(z)$ the sought pdf we write

$$f_Z(z) = P(in\ X)\frac{1-F_X(z)}{m_X} + P(in\ Y)\frac{1-F_Y(z)}{m_Y}$$

where $P(in\ X)$ and $P(in\ Y)$ are those evaluated above. We finally get

$$f_Z(z) = \frac{2 - F_X(z) - F_Y(z))}{m_X + m_Y}$$

P.2.7 (2.40) Arrivals Occur in a periodic way, alternating constant intervals of value of T and $2T$. Taken a Random Inspection Point, find

a) the probability that the RIP lies in an interval of length $2T$;

b) the pdf of the waiting time to the next arrival;

c) the probability an arrival occurs in the next Δt;

Solution
a) 2/3

b) The pdf of the waiting time is given by the Total Probability Theorem by distinguishing the possible intervals. We have

$$f_Y(y) = \frac{1}{3}\mathrm{rect}_T(x) + \frac{2}{3}\mathrm{rect}_{2T}(x)$$

where $rect_D(x)$ is the pdf that is uniform, with value $1/D$ in $[0; D]$.

c) The sought probability is

$$f_Y(0)\Delta t = \frac{1}{3T} + \frac{2}{6T} = \frac{2}{3T}$$

This is also the asymptotic rate times Δt. The rate can also be evaluated as the reverse of the average interarrival time $(3/2)T$.

P.2.8 (2.41) Arrivals occur according to Renewal Events where the interarrival period has pdf Erlang-k, $k > 1$ with parameter λ. Assuming that the first arrival is in the origin (t=0), find $P(A(t, t + \Delta t) = 1)$, disregarding infinitesimals $(\Delta t)^k$ of order $k > 1$, in the two cases

a) $t = 0$;

b) $t \to \infty$.

and compare with the case $k = 1$.

c) Find the probability that the first arrival after the origin occurs in $[t, t + \Delta t]$, in the two cases $k = 1, 2$.

Solution
Pdf Erlang-k writes

$$f_X(x) = \frac{(\lambda x)^{k-1}}{(k-1)!}\,\lambda\, e^{-\lambda x} \qquad\qquad x \geq 0$$

a) The sought probability is

$$P(0, 0 + \Delta t) = f_X(0)\Delta t = \begin{cases} 0, & k > 1 \\ \lambda\Delta t, & k = 1 \end{cases}$$

Questa vale 0 per $k > 1$, e $\lambda\Delta t$ per $k = 1$

b) is the reverse of the average interarrival time $\times\ \Delta t$, i.e., $1/(\lambda k)\Delta t/T$.

c) It is exactly the probability that the interarrival time equals t, i.e., by the definition of pdf:

$$f_X(t)\Delta t = \frac{(\lambda t)^{k-1}}{(k-1)!}\,\lambda\, e^{-\lambda t}\Delta t$$

P.2.9 Given Renewal Arrivals of rate λ, find the probability $P(A(t,t+\Delta t)=1, A(t+\Delta t, t+2\Delta t)=1)$ where t is a Random inspection point and the pdf $f_X(x)$ of the interarrival time is

a) Negative exponential

b) Erlang-2.

b) Uniform, starting from 0.

Note that the parameters of pdfs' above can be derived by the rate of the arrivals λ

Solution - a) Due to the independence property of Poisson Arrivals we have

$$P(A(t,t+\Delta t)=1, A(t+\Delta t, t+2\Delta t)=1) = \lambda^2(\Delta t)^2$$

b) Since an Erlang-2 with parameter μ has average $2/\mu$, the parameter of our distribution is $\lambda/2$, and its average is $1/\lambda$. According to the first Renewal Theorem we have

$$P(A(t,t+\Delta t)=1) = \lambda\Delta t$$

and,

$$P(A(t+\Delta t, t+2\Delta t)=1|A(t,t+\Delta t)=1) = f_X(0) = 0\Delta t$$

Therefore

$$P(A(t,t+\Delta t)=1, A(t+\Delta t, t+2\Delta t)=1) = 0$$

c) The rightmost extreme of the uniform pdf is $T=2/\lambda$. Then, the evaluation proceeds as in point b)

$$P(A(t,t+\Delta t)=1) = \lambda\Delta t$$

and,

$$P(A(t+\Delta t, t+2\Delta t)=1|A(t,t+\Delta t)=1) = f_X(0) = (\lambda/2)\Delta t$$

Therefore

$$P(A(t,t+\Delta t)=1, A(t+\Delta t, t+2\Delta t)=1) = (\lambda^2/2)(\Delta t)^2$$

P.2.10 The same as the previous Problem where now the probability to be found is $P(A(t,t+\Delta t)=0, A(t+\Delta t, t+2\Delta t)=1)$

P.2.11 (3.22) Poisson arrivals are deleted according to the following procedure: arrival i is suppressed with probability p if arrival $i-1$ has been suppressed, otherwise it is suppressed with probability q.

a) Find the asymptotic probability ($i \to \infty$) of suppression of an arrival (Hint: Consider the discrete-time binary process S_i *1 non-suppression, 0 suppression.* This process carries memory, ...);

b) Taken a survived arrival i, find the probability P_k that point $i+k$ is the first survived after i (the answer does not depend on the answer to point a); take at first $k=1$ and then...);

c) find the average distance between survived arrivals.

Solution - a) Process S_i is binary MC with transition probabilities

$$p_{11} = 1-q \qquad p_{10} = q \qquad p_{01} = 1-p \qquad p_{00} = p$$

The asymptotic distribution is

$$\pi_1 = \frac{1-p}{1-p+q}$$

b)

$$P_1 = 1-q \qquad P_k = qp^{k-2}(1-p) \qquad k \geq 2$$

c) the sought distance is the reverse of the arrival rate of survivors $\lambda P(N)$:

$$m = \frac{1}{\lambda}\,\frac{1-p+q}{1-p}$$

Chapter 3

Elements of Traffic Theory

3.1 Unlimited resources

3.1.1 Single-user Traffic and Work

The *Traffic* is an activity originated by a *User* that requires the availability of a *resource* that we call *server*. In this section we assume that the server is always available whenever the user becomes active. In this case the traffic generated by a user is represented by a binary process $Z(t), (Z = 0, 1)$ that represents the activity of the user/server, $Z = 1$ representing the *active/busy state*, and $Z = 0$ representing the *idle state* (Figure 3.1). In stationary conditions the average traffic is

$$S = E[Z(t)] = \pi_1. \tag{3.1}$$

When the process is ergodic, denoted by t_1 the time in $[0; t]$ the process is in state 1, we have

$$S = \pi_1 = \lim_{t\to\infty} \frac{t_1}{t}. \tag{3.2}$$

The traffic is a pure number, but we say it is measured in *Erlang*.

Here we consider positive-recurrent regenerative traffic, or ergodic, in stationary conditions, where regeneration instants occur when the busy period starts, the duration of the idle period is represented by the RV Y, with average m_Y and the duration of the activity period is represented by RV and X of m_X. From the basic results (1.196) we have

$$S = E[Z] = \pi_1 = \frac{m_X}{m_X + m_Y}. \tag{3.3}$$

Often S is simply called traffic, implying the mean of the stationary binary traffic process. Also, activation instants are also called *service requests* or simply *arrivals*.

The frequency of activation instants occurs at periods (cycles) with average $m_Y + m_X$ and therefore, by Theorem 2.24 the frequency of activation instants is

$$\lambda = \frac{1}{m_Y + m_X}, \tag{3.4}$$

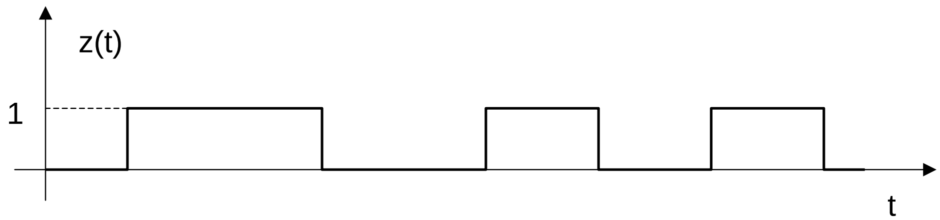


Figure 3.1: *Traffic sample of a single user.*

and we have

$$S = \lambda m_X. \tag{3.5}$$

The area of the traffic function at time t

$$W(t) = \int_0^t Z(t)dt, \tag{3.6}$$

is called the *Work* of the user at time t, measured for example in sec, bit, bytes. By (**??**) and (3.2) it appears also that asymptotically the traffic represent the Work per unit time generated by the user, in Erlang units, but also in bit/sec, or byte/sec.

3.1.2 Multi-user traffic

Within transmission equipments the traffic from various users is merged (multiplexed) and the total traffic process is the sum of the single-user traffic processes, $Z(t) = \sum Z_i(t)$. Then the traffic is a multi-state process representing the number of users that are simultaneously active.

When the component traffics $Z_i(t)$ are independent, as we will always assume, and have the same average value, the cumulative traffic $Z(t)$ presents the well known binomial distribution. In fact, the probabilityof having k active users out of n equals the probability of success in n independent trials with success probability equal to $\frac{m_X}{m_Y+m_X}$. Therefore, the distribution is

$$P(Z = k) = \binom{n}{k} \left(\frac{m_X}{m_Y + m_X}\right)^k \left(\frac{m_Y}{m_Y + m_X}\right)^{n-k}, \qquad 0 \le k \le n. \tag{3.7}$$

The average is the sum of the averages:

$$S = n\frac{m_X}{m_Y + m_X}. \tag{3.8}$$

Even the "arrival" fluxes merge, and the global flux has frequency

$$\lambda = n\frac{1}{m_Y + m_X},$$

and, therefore, we still have

$$S = \lambda m_X, \tag{3.9}$$

which generalizes the (3.5).

We remark the fact that (3.8) holds whatever the pdfs of the service and idle times X and Y.

Also in the multi-user case the Work is the area of the traffic curve, and the global Work is the summation of the Works of the single users.

3.1.3 Infinite population model

We already saw in Theorem 2.40 that if we have a very large number n of users, and their average idle period m_Y grows as n in such a way that the global arrival frequency is λ, then, in the limit $n \to \infty$ the global arrival process becomes Poisson. Furthermore, when the activity time has average m_X, for the traffic distribution, whose average we denote by S, can be derived from (3.7), for $n \to \infty$ e $m_Y = kn$, and $m_Y = kn$, and tends to the Poisson distribution. (the demonstration is similar to that already seen in section 1.3.1). We have thus proved the

Theorem: (3.10)
The traffic with the infinite population model and frequency λ, with any pdf of the activity time of average m_X, is Poisson distributed.

$$P(Z = k) = \frac{(\lambda m_X)^k}{k!} e^{-\lambda m_X}, \tag{3.11}$$

The average of traffic (3.11) is still

$$S = \lambda m_X. \tag{3.12}$$

We remark the fact that (3.11) holds whatever the pdf of the service time X.

When arrivals can not be modeled by a Poisson process, then (3.11) no longer holds. However (3.12) is shown to hold even in this case by the following

Theorem: *Little's Result* (3.13)
For any ergodic traffic in stationary conditions we have

$$S = \lambda m_X.$$

Proof
We assume that $Z(t)$ is a regenerative positive-recurrent process, where regeneration points are represented by the starting point of the busy period. We start from (1.204):

$$S = E[Z(t)] = \frac{E\left[\int_C Z(t)dt\right]}{E[C]}. \tag{3.14}$$

We also have

$$\int_C Z(t)dt = \sum_{k=1}^{A} X_k, \tag{3.15}$$

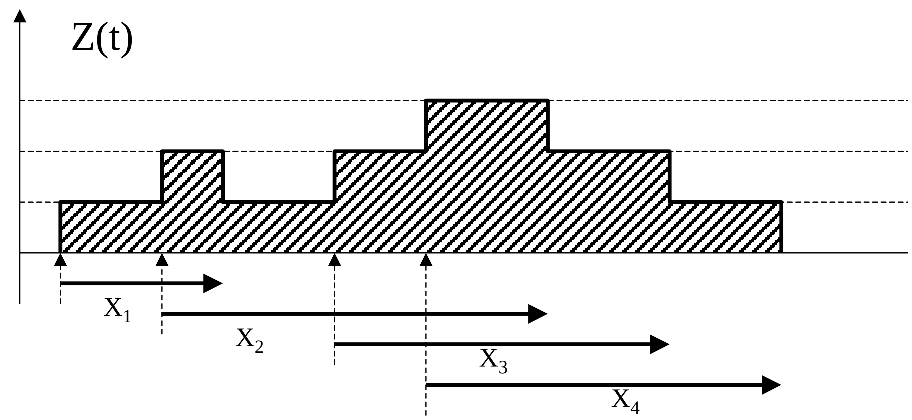


Figure 3.2: *The area equals the sum of X_i.*

where A is the RV that represents the arrivals within a regeneration cycle. The (3.15) is always true for any sample of the process, and can be easily derived observing Figure 3.2. By substituting (3.15) into (3.14), we get

$$S = \frac{E\left[\int_C Z(t)dt\right]}{E[C]} = \frac{E\left[\sum_{k=1}^{A} X_k\right]}{E[C]} = \frac{E[A]m_X}{E[C]} = \lambda m_X.$$

We have used equality

$$\frac{E[A]}{E[C]} = \lambda,$$

that is (3.14) applied to the arrival process which is a discrete-time process, where the integral becomes a summation.

A similar demonstration, replacing operations over the cycle with averages over the entire time axis, can be applied to ergodic traffic.♣

Once more, Little's result extends to stochastic processes a property that holds for deterministic periodic samples.

Example (3.16)
In an office, each employee generates an average traffic of 0.25 erlang. If there are 10 employees

- *a) what is the average traffic generated?*
- *b) what is the probability that five or more employees are calling at the same time?*
- *c) If a call lasts 3 minutes in the average, how many new calls/minute are generated on the average?*

a) The average traffic is the sum of the average traffics. So the average global traffic is 2.5 Erlang.
b) The probability that an employee is making a call is $p = 0.25$. So the probability of having 5 or more employees calling is

$$P(N \geq 5) = \sum_{k=5}^{10} \binom{10}{k} p^k (1-p)^{10-k} = 0.068.$$

c) From the Little's Result we have $\lambda = S/m_X = 2.5/3 = 0.833$ calls/minute.♣

3.2 Limited resources

In the preceding sections we have implicitly assumed that service is immediately provided to any user that activates, meaning that any user can immediately make use of the resources it needs, or, again, that resources are unlimited. In practice, resources are limited, and this imply that the needed resource can be unavailable. Here, the activation of sources is seen as arrivals of customers to a Service Systems, and their de-activation as departures of the customers from the Service System.

A limited number of servers may either cause the user activation to be *blocked*, and its traffic and work deleted, or to be *delayed*, in which case the work is done later in time. In the latter case arrivals are accommodated into a *queue* where customers await their turn for service. In an other case again, service is shared among users, so that the service speed is reduced and the service time is increased, in a dynamic way as the number of users in service changes. In the latter cases, also, some blocking can be introduced, for example when the queueing room is finite.

In the light of what said above, arrivals are seen as carrying an *offered traffic* (or *offered work*) which refer to the case of unlimited resources. On the other side the *served traffic* refer to the users that are not blocked and get service. We denote by $N(t)$ the process *number of users in the system at time t*, that we call *occupancy process*, whose asymptotic distribution elements we denote as π_i. Process $N(t)$ coincide with the served traffic if blocking used, and, otherwise, can always be decomposed as $N(t) = N_s(t) + N_c(t)$, the sum of traffic and the number of users in the queue. Next we see some general results.

3.2.1 Flow balance

Let refer to arrivals, one at a time, with stationary rate λ_a and departures, again, one at a time, with stationary rate λ_d. In a similar way let λ_i be the conditional stationary arrival rate when $N(t) = i$, and μ_j be the conditional stationary departure rate when $N(t) = j$.

As we have seen in Section 1.7, that balance equations hold for any process in stationary conditions, even for the process considered, which is a general form of birth and death process, though not markovian in general. Therefore we can write, as in the markovian B&D process,

$$\pi_i \lambda_i = \pi_{i+1} \mu_{i+1}. \tag{3.17}$$

Taking the summation on the whole state space we get

$$\lambda_a = \sum_i \pi_i \lambda_i, \qquad \lambda_d = \sum_i \pi_{i+1} \mu_{i+1},$$

and because of (3.17) we finally have

$$\lambda_d = \lambda_a, \tag{3.18}$$

a result somehow obvious, another expression of balance equations that states that in stationary conditions the input and output user flows must balance. This result can be easily extended to the case with multiple arrivals and departures, resulting in the

Theorem: (3.19)
In stationary conditions, arrivals and departures at a system presents the same rate.

3.2.2 Little's Result

Another relevant process, defined over the discrete axis i, $i = 0, 1, 2, \ldots$, is the time V_i spent by user i in the system, while getting service, or waiting, or doing else; in other words the time between users' arrival and departure. If we look at a sample of $N(t)$, it can be treated exactly as we did with traffic in Figure 3.2. In the same way we can prove again the Little's Result 3.13, where the activation frequency is now replaced with the arrival frequency (we use the same symbol), an the activity time is now replaced by the time in the system V_i. Therefore we have

$$E[N] = \lambda E[V]. \tag{3.20}$$

3.2.3 PASTA property

Another relevant process is the discrete-parameter process $N'_h = N(t_h^-)$, i.e., the sampling of $N(t)$ at time t_h^- defined as the time of arrival of user h just before his entrance into the system (Figure 3.3). Note that this process is defined over the discrete h axis.

In general, processes $N(t)$ and N'_h have different asymptotic distributions. The asymptotic distribution of $N(t)$ is the one that is seen by a RIP, a random inspection point. The asymptotic distribution of N'_h is the one that is seen by a RIP on h axis, i.e., the number of a user selected at random. However, some arrival processes are such that the stationary distributions of $N(t)$ and N'_h coincides. Those are called ASTA processes, where the acronym stands for Arrivals that See Time Averages; term Time Averages is used as synonym of means and distributions, since we usually refer to Ergodic processes. We now prove the PASTA property.

Let denote by π_i and q_i the stationary distributions of $N(t)$ and N'_h respectively, while λ denotes the arrival frequency and $A(t, t+\Delta t)$ the arrival in $t, t+\Delta t$.

Theorem: (3.21)
In stationary conditions Poisson arrivals sees time averages, i.e., $q_i = \pi_i, \forall i$.

We have

$$\begin{aligned} q_i &= P(N'_h = i) = P(N(t) = i | A(t, t+\Delta t)) = P(A(t, t+\Delta t) | N(t) = i)\ \frac{P(N(t) = i)}{P(A(t, t+\Delta t))} = \\ &= P(A(t, t+\Delta t) | N(t) = i)\ \frac{\pi_i}{\lambda \Delta t}. \end{aligned} \tag{3.22}$$

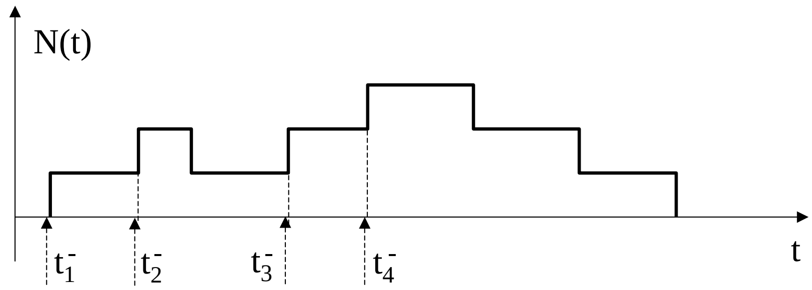


Figure 3.3: *Samples of the occupancy process $N(t)$ at arrival instants.*

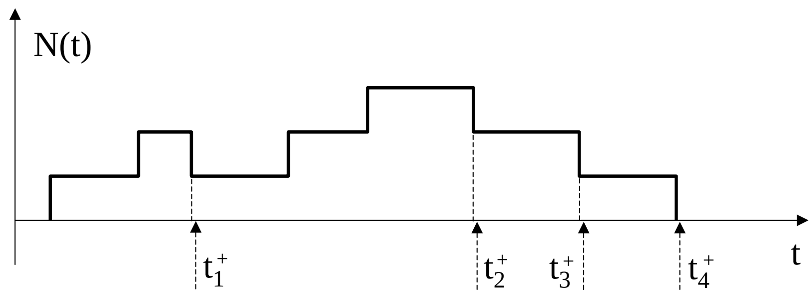


Figure 3.4: *Samples of the occupancy process $N(t)$ at departure instants.*

$N(t)$ is completely determined by the whole arrival history, $H(t)$, up to time t. On the other side, $A(t, t+\Delta t)$ does not depend on such history by definition, and we have

$$P(A(t,t+\Delta t)|N(t)=i) = P(A(t,t+\Delta t)) = \lambda \Delta t,$$

which substituted into (3.22) proves the thesis.♣

If arrivals are not Poisson, then we have another interesting property. As we did with arrivals, let r_i denote the the stationary distribution of $N''_k = N(t_k^+)$, the sampling of $N(t)$ just after the exit of user k (Figure 3.4). We have

Theorem: (3.23)
In stationary conditions , the distribution q_i seen at the entrance and the distribution r_i seen at the exit can be expressed as

$$q_i = \pi_i \frac{\lambda_i}{\lambda}, \tag{3.24}$$

$$r_i = \frac{\mu_{i+1}}{\bar{\lambda}}\ \pi_{i+1}, \tag{3.25}$$

and are equal, $q_i = r_i$.

In fact, (3.24) comes from (3.22). In a similar way, defining by $D(t, t + \Delta t)$ the event departure in $[t + \Delta t]$, we can write

$$r_i = P(N''_k = i) = P(N(t) = i+1|D(t,t+\Delta t)) = P(D(t,t+\Delta t)|N(t) = i+1)\ \frac{P(N(t) = i+1)}{P(D(t,t+\Delta t))}.$$

In stationary conditions we have

$$P(N(t) = i + 1) = \pi_{i+1},$$
$$P(D(t, t + \Delta t)) = P(A(t, t + \Delta t)) = \lambda \Delta t,$$
$$P(D(t, t + \Delta t)|N(t) = i + 1) = \mu_{i+1}\Delta t,$$

and, therefore, substituting we get (3.25). The equality of the distributions descends recognizing that (3.25) equals (3.24) through (3.17). ♣

3.3 Markovian Blocking Systems

Here we consider the markovian case where users arrive according to a markovian process, i.e. with either the finite or infinite population model, and service times have negative exponential pdf with rate μ, and that the number of servers is m, while possibly the contemporary requests are unlimited. We must also specify what happens when a user wants to be active but there are no resources left.

In telephone networks, the model used is the one called *calls cleared,*. This means that when a request can not be satisfied the user/user is cleared, that is, it returns to the idle state as if it ended its activity period. This is what is known as *Blocking System*.

3.3.1 Infinite Population Blocking System

Here we refer to the Infinite Population model, where arrivals occur according to the Poisson Model at rate λ. We assume that service times have negative exponential pdf with rate μ, and that the number of servers is m. With these assumptions the number of active users $N(t)$ is the Birth and Death Process with parameters

$$\lambda_i = \lambda, \qquad \mu_i = i\mu$$

whose solution is the one already seen in (1.154) for the generalized random walk

$$\pi_i = \pi_0 \frac{\lambda_0\ \lambda_1\ \lambda_2 \ldots \lambda_{i-1}}{\mu_1\ \mu_2 \ldots \mu_i} = \pi_0 \frac{\lambda^i}{i!\mu^i}, \qquad i = 1, \ldots, m. \tag{3.26}$$

Using traffic $S = \lambda/\mu$ and normalizing we get

$$\pi_i = \frac{S^i/i!}{\sum_{k=0}^{m} S^k/k!}, \qquad i = 1, \ldots, m. \tag{3.27}$$

The distribution above is the truncated Poisson, also known as the Erlang distribution. Note that as $m \to \infty$ the results coincides with the distribution (3.11), which has been derived assuming that any user that wants to become active can do so, since it always can find the required resource. Note also that (3.27) coincides with (3.11) conditioned to $i \leq m$ (this could be suspected, but is not at all obvious).

When a customer arrives, because of the PASTA Theorem (3.21) it sees the stationary distribution above. Therefore it is rejected if arrives when the system is in state m (congestion), which happens with probability

$$\pi_m = \frac{S^m/m!}{\sum_{k=0}^{m} S^k/k!} = B_m(S). \tag{3.28}$$

This is called Erlang B formula. Evaluating this formula for high m and a, as it happens in telephone networks, is rather difficult. However, we have a recursion that simplifies the problem:

$$B_m(S) = \frac{S\ B_{m-1}(S)}{m + S\ B_{m-1}(S)} \tag{3.29}$$

$B_m(S)$ is function that decreases with m and increases with S. However, if we keep the traffic per server $\rho = S/m$ (also called the server load factor) constant, then $B_m(\rho m)$ decreases as m increases, and goes to zero when $m \to \infty$ and $\rho < 1$. Setting $\rho = 1 - \varepsilon \simeq 1$, the latter result shows that as $m \to \infty$ we get zero blocking with a number of servers that is equal to the offered traffic. This, of course, is the effect of the law of Large numbers, that reduces the traffic oscillations about its average as the average goes to infinity.

On the other side, if we keep $B_m(S)$ constant, then ρ increases as S increases. This shows the effect of large numbers: in order to have great efficiency and low Blocking probability it is convenient to merge traffics. This is clearly shown in Figures 3.5 and 3.6.

Figure 3.5 shows the blocking probability as function of the number of resources (lines, phones) available for different values of the offered traffic S. We see that to get a blocking probability equal to 0.01 with $S = 30$ we need a number of resources equal to 40. With the same blocking and $S = 90$ the needed resources are 107, with a gain of 13 compared to the case where the 90 Erlang were served in three disjoint groups of 30 resource each. This show the convenience of grouping the resources.

In Figure 3.6 we see the strong gain in efficiency that occurs in going from 1 to 100 resources. Note the case in which the offered traffic/server exceeds 1. Here the traffic that exceeds one is rejected.

The success of Erlang B formula is also due to the following property

Property *Insensitivity to the service-time pdf:* (3.30)

The Erlang B formula holds true also when the service time is no longer negative exponential, but has a general distribution, as it happens for (3.11).

The demonstration is rather involved and is omitted.

Finally, we have to remark that the frequency of the blocked and served traffic is $B\lambda$, and $(1-B)\lambda$ respectively; the blocked and served traffic are BS and $(1-B)S$, the latter being also the average number of users (busy servers) in the system. Furthermore we have

Property (3.31)

blocking is not an independent process and, therefore, the flow of blocked and served users is no longer Poisson.

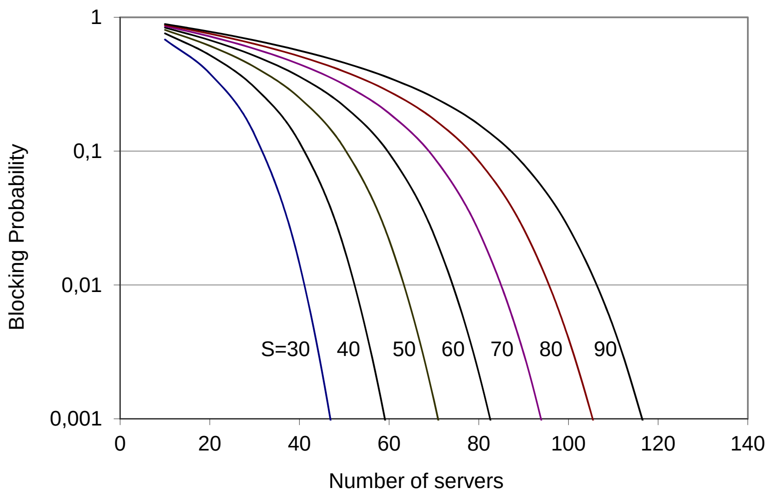


Figure 3.5: *Blocking probability as function of the number of servers m for various values of offered traffic.*

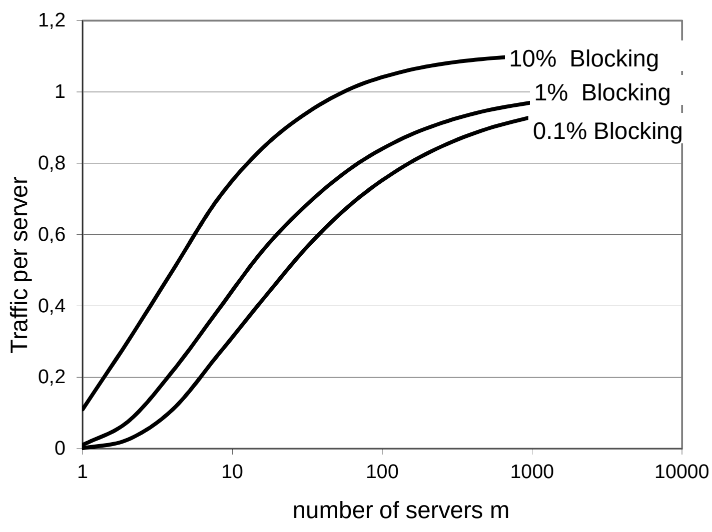


Figure 3.6: *Utilization of servers (traffic/servers) as function of the number of servers for different blocking probabilities.*

The property can be proved showing that $P(B2|B1)$, i.e., the probability that, given that user 1 is blocked, also user 2, arriving next, is blocked, is not equal to $P(B2)$. After arrival 1 is blocked, the first free server occurs after a negative exponential time of rate $m\mu$, while the next arrival occurs after a negative exponential time of rate λ. User 2 is blocked if the first RV is less than the second, whose probability is

$$P(B2|B1) = \frac{\lambda}{\lambda + m\mu} \neq P(B2)$$

3.3.2 Finite Population Blocking System

Here we refer to the Finite Population with M users and m servers, $m \leq M$, where each user originates a traffic that is a binary MC (a special case of what seen in Section 3.1.2). Service request are originated after a negative exponential Idle Period with frequency $\nu = 1/m_Y$, whereas service times have negative exponential pdf with rate μ. With these assumptions the number of active users $N(t)$ is the B&D Process with frequencies

$$\lambda_i = (M - i)\ \nu, \tag{3.32}$$

$$\mu_i = i\ \mu \qquad i \leq m$$

The solution is (3.26):

$$\pi_i = \pi_0 \frac{\lambda_0\ \lambda_1\ \lambda_2 \ldots \lambda_{i-1}}{\mu_1\ \mu_2 \ldots \mu_i} = \pi_0 \frac{M(M-1)\ldots(M-i+1)\nu^i}{i!\mu^i} \qquad i \geq 0 \tag{3.33}$$

and setting $\alpha = \nu/\mu$, we have

$$\pi_i = \pi_0 \binom{M}{i} \alpha^i, \qquad 0 \leq i \leq m$$

$$\pi_i = \frac{\binom{M}{i} \alpha^i}{\sum_{k=0}^{m} \binom{M}{k} \alpha^k}, \qquad 0 \leq i \leq m. \tag{3.34}$$

This is called the Truncated Binomial or Engset distribution, and when $M = m$ it becomes the binomial

$$\pi_i = \binom{M}{i} \left(\frac{\alpha}{1+\alpha}\right)^i \left(\frac{1}{1+\alpha}\right)^{M-i}$$

where $\alpha/(1+\alpha)$ coincides with the offered traffic of a single user $m_X/(m_X + m_Y) = \nu/(\nu + \mu)$, already seen in (3.7).

Probability π_m is the probability that the system is full in stationary conditions. This is called *time congestion*, because in ergodic systems the probability coincides with the fraction of time the event occurs. However, in this case arrivals are no longer Poisson, and the distribution seen by users does not coincide with the continuous-time distribution.

From 3.23 , and using (3.32), we have

$$q_i = \frac{(M-i)\nu}{\lambda}\, \pi_0 \binom{M}{i} \alpha^i = \frac{(M-i)}{M}\, q_0 \binom{M}{i} \alpha^i = q_0 \binom{M-1}{i} \alpha^i, \qquad 0 \le i \le m,$$

where we have used

$$q_0 = \frac{M\nu}{\lambda}\, \pi_0.$$

After normalizing we get

$$q_i = \frac{\binom{M-1}{i} \alpha^i}{\sum_{k=0}^{m} \binom{M-1}{k} \alpha^k} \qquad i = 0, 1, \ldots, m. \tag{3.35}$$

Note that this coincides with π_i when the population size is $M-1$. This is not an accident, as the arriving user can not see itself. The blocking probability is now

$$q_m = \frac{\binom{M-1}{m} \alpha^m}{\sum_{k=0}^{m} \binom{M-1}{k} \alpha^k} \tag{3.36}$$

This is called *call congestion*, since it represents the fraction of users (calls) that find the system full, and the corresponding formula is named the Engset formula. This formula has properties similar to the Erlang B, and actually tends to the Erlang B when the population size M increases. It can also be expressed as function of the offered traffic S by using the substitution $\alpha = S/(M-S)$.

Property *Insensitivity to the service-time pdf:* (3.37)
Distributions (3.34) and (3.35) can be shown to hold for service times with general pdf, although Idle Period must remain negative exponential, otherwise the above formulas do not hold.

3.4 Queueing systems

3.4.1 Ergodicity

When $N(t)$ is ergodic and no blocking exists, the flow balance guarantees that the users' service rate (departure rate) is equal to the arrival rate. This also means that the *serviced traffic* is equal to the *offered traffic*, that is, the traffic curve is altered by the delay (this process is exemplified in Figure 3.7 with $M = 2$) and, since all the work is served, the area of the traffic curve does not change in the reshaping of the traffic curve. It can happen however, that the offered traffic is greater than the service capability. In this case, the work builds up in the queue, $N(t)$ increases beyond any limit and is not ergodic. We refer to this condition as *instability.*

Under very general assumptions, provided that the arrival process and the service process are stationary, the following theorem holds

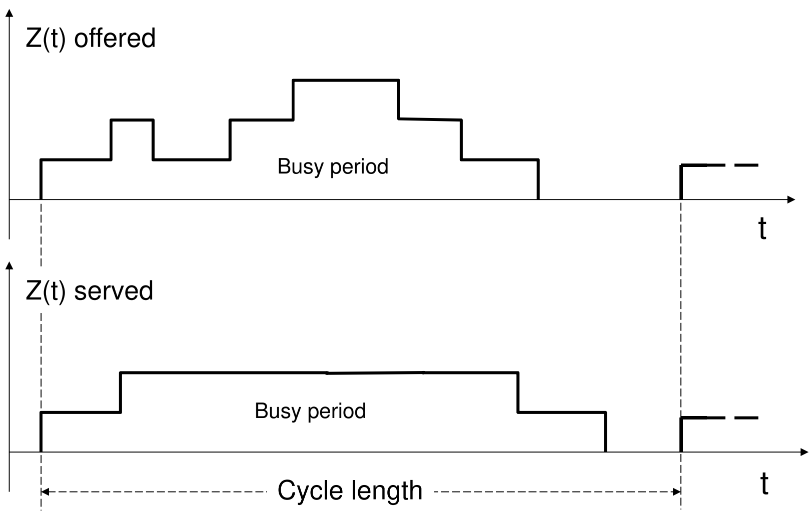


Figure 3.7: *The area equals the sum of X_i.*

Theorem: (3.38)
In a queue with m servers the system content $N(t)$ is ergodic if $\lambda m_X < m$, and is not ergodic if $\lambda m_X > m$.

We do not provide the detailed proof of the theorem, which requires some technicalities. However, the theorem itself is very intuitive. In fact, if $\lambda m_X < m$, $N(t)$ can not increase with time, otherwise all servers become busy forever, servicing a work equal to mt for each t, while in the same time the average work entering the system is $\lambda m_X t < mt$, which is a contradiction. On the other side, when $\lambda m_x > m$, the work entering the system is greater than the servers can manage, and the work keeps cumulating in the queue.

3.5 Markovian Queueing Systems

When all servers are busy, an alternative to blocking is queueing, i.e., users that can not immediately be served wait in line and reach a server whenever one becomes free. Here we are interested to different processes, such as the number $N(t)$ of users in the system, or the number $N_q(t)$ of users in the queue, or, again, the number $N_s(t)$ of users being served (actually the traffic served).

We have seen about Theorem 3.38 that queueing changes the traffic process shape but the work is entirely conserved if $S < m$. This condition can be written as

$$\rho = \frac{S}{m} = \frac{\lambda m_X}{m} < 1,$$

where ρ, thraffic per server, is also called the server *load factor* and, of course, can not be greater than one. In fact, it also represents the asymptotic (stationary) probability that the server is busy, or the average number in the server itself.

3.5.1 M/M/1/∞ System

M/M/1 systems are characterized by Poisson arrivals with rate λ and one server, with independent negative exponential service times.

Again, $N(t)$ is a Markov Birth and Death process with $\lambda_i = \lambda$ e $\mu_i = \mu$, that coincides with the random walk in continuous time. The asymptotic distribution exists for $\lambda < \mu$ and is the geometric

$$\pi_i = (1-\rho)\rho^i \qquad\qquad i = 0, 1, \ldots$$

Where $\rho = \lambda/\mu$ is the load factor, and must be smaller than one in order the solution above to hold. Note that $\pi_0 = 1-\rho$. This comes also from the fact that with probability $1-\rho$ the server is idle, and, with just one server, when the server is idle the system is empty. Taking the average of the distribution above we have

$$E[N] = \frac{\rho}{1-\rho}$$

and, by the Little's Result, we get the average time through the system

$$E[V] = \frac{m_x}{1-\rho} = \frac{1}{\mu}\,\frac{1}{1-\rho}.$$

For the time spent in the queue we have $E[W] = E[V] - m_x$, and therefore

$$E[W] = m_x\frac{\rho}{1-\rho} = \frac{1}{\mu}\,\frac{\rho}{1-\rho} \tag{3.39}$$

It should be noted that all the average times above are proportional to the average service time $m_X = 1/\mu$. This implies that using a server k times more powerful, and increasing by k also the arrival rate, so to maintain the same load ρ, reduces by k times all the average times (even though the average number in the system is the same).

For $\rho \geq 1$ the solution does not exists, meaning that the chain is not positive-recurrent, the number and the time in the system increase without bounds, each user getting zero service for an infinite amount of time and never departing.

Example (3.40)
Find the change of $E[N], E[V], E[W]$ when m M/M/1 queues are merged into one M/M/1 queue with a server m times faster.

For each of the queues we have

$$E[N] = \frac{\rho}{1-\rho}, \qquad E[V] = \frac{1}{\mu}\,\frac{1}{1-\rho}, \qquad E[W] = \frac{1}{\mu}\,\frac{\rho}{1-\rho},$$

and the global average number of users within the m systems are $mE[n]$.

In the merged system the load factor does not change $(m\lambda)/(m\mu) = \lambda/\mu$. Therefore we have

$$E[N] = \frac{\rho}{1-\rho}, \qquad E[V] = \frac{1}{m\mu}\,\frac{1}{1-\rho}, \qquad E[W] = \frac{1}{m\mu}\,\frac{\rho}{1-\rho}.$$

We see that all figure are divide by m with respect to the original systems. Again, we see it is convenient to merge resources.♣

We see now the pdfs' of the time in the system V and the waiting time $W = V - X$.

Theorem: (3.41)
The time in the system is a negative exponential RV with average

$$E[V] = \frac{1}{\mu}\,\frac{1}{1-\rho}$$

Proof
When a user arrives and finds i users before him, he has to wait for a time hat has Erlang-i pdf, equal to the sum of the service time X_i of the user in the service (here the remaining service time is still exponential) and the service times of the $i-1$ users in the queue. Then, we must add the service time of the arriving user. The, thanks to the Total Probability Theorem, we have

$$f_V(x) = \sum_{i=0}^{\infty} q_i E_{i+1}(x)$$

where q_i is the occupancy distribution seen by an arrival, that in this system is equal to π_i. By expliciting we have

$$\begin{aligned} f_V(x) &= \sum_{i=0}^{\infty}(1-\rho)\,\rho^i \frac{(\mu x)^i}{i!}\mu\, e^{-\mu x} = \mu(1-\rho)e^{-\mu x}\sum_{i=0}^{\infty}\frac{(\mu\rho x)^i}{i!} = \\ &= \mu(1-\rho)e^{-\mu x}\, e^{\mu\rho x} = \mu(1-\rho)e^{-\mu(1-\rho)x} \qquad x \geq 0. \clubsuit \end{aligned}$$

In a similar way we can show the following

Theorem: (3.42)
the conditional waiting time pdf is

$$f_W(x|W>0) = \mu(1-\rho)e^{-\mu(1-\rho)x} \tag{3.43}$$

while the unconditional one is

$$f_W(x) = (1-\rho)\delta(x) + \rho f_W(x|W>0),$$

whose average is exactly the (3.39). ♣

Here also the size of the system may be finite, so that the maximum number of users that can be accommodated in the system is Q. If users that arrive when the system is full leave with no consequence, the chain becomes truncated and the solution is the same but the normalization constant (truncated geometric distribution):

$$\pi_i = \frac{1-\rho}{1-\rho^{Q+1}}\,\rho^i \qquad i = 0, 1, \ldots, Q$$

and exists for any ρ, which is now called *offered* load factor. The frequency of the served arrivals is now $\lambda_s = \lambda(1 - \pi_Q)$, and the served traffic is now

$$\rho_s = \rho(1 - \pi_Q) = \rho \frac{(1 - \rho^Q)}{1 - \rho^{Q+1}} = 1 - \pi_0. \tag{3.44}$$

Served traffic and workload

The result in (3.44) shows that the served traffic equals $1 - \pi_0$. This is a result of wider relevance and must be commented.

In the definition of traffic and work we have seen that delaying changes the shape of the traffic curve, i.e., the number of served users, but not the workload and the traffic; for example when there is only one server the server traffic curve becomes binary $(0, 1)$; the original multi-level traffic curve is squeezed and lengthened to become a zero/one sample function, however the workload, and the traffic value, are maintained. Therefore, if we refer to the traffic of the single server in ergodic conditions, the served traffic is, according to the definition, the probability that the server is active, i.e., $1 - \pi_0$. This is equal to the offered one if no traffic is lost. If some traffic is lost, for example by blocking, then $1 - \pi_0$ represents the served traffic, which is the offered traffic minus the blocked one.

Note that in the argument above we have not made use of the markovian assumption. Then the property above is general and holds true even for non-markovian single-server systems.

3.5.2 M/M/m/∞ system

Here the number of servers, all of the same type, is increased to m. It is again a Birth and Death process with

$$\lambda_i = \lambda$$

$$\mu_i = \begin{cases} i\mu & i \le m \\ m\mu & i \ge m \end{cases}$$

Denoting, as usual, the server load factor as $\rho = \lambda/(m\mu)$, the solution is

$$\pi_i = \begin{cases} \pi_0 \dfrac{(m\rho)^i}{i!} & i \le m \\[2ex] \pi_0 \dfrac{1}{m!} \dfrac{(m\rho)^i}{m^{i-m}} & i \ge m. \end{cases} \tag{3.45}$$

In order the distribution to converge we must have $\rho < 1$, in acordance with Theorem 3.38. We observe that

$$C_m = \sum_{i=m}^{\infty} \pi_i = \pi_0 \frac{(m\rho)^m}{m!} \sum_{k=0}^{\infty} \frac{(m\rho)^k}{m^k} = \pi_0 \frac{(m\rho)^m}{m!} \frac{1}{1 - \rho} \tag{3.46}$$

which provides

$$\pi_0 = \left(\sum_{k=0}^{m-1} \frac{(m\rho)^k}{k!} + \frac{(m\rho)^m}{m!} \frac{1}{1-\rho} \right)^{-1} .$$

The (3.46) represents the probability of waiting, i.e., the probability that all servers are taken, and explicitly can be written, as function of the traffic $S = m\rho$, as

$$C_m(S) = \frac{S^m/m!}{(1 - \frac{S}{m}) \sum_{k=0}^{m-1} S^k/k! + S^m/m!}, \tag{3.47}$$

and is known as Erlang C.

The average number in the system can be evaluated as

$$\begin{aligned} E[N] &= \sum_{i=1}^{\infty} i\pi_i = \pi_0 \sum_{i=1}^{m} i \frac{m^i}{i!} \rho^i + \pi_0 \frac{m^m}{m!} \sum_{i=m+1}^{\infty} i\rho^i = \\ &= \pi_0 m\rho \sum_{i=1}^{m} \frac{m^{i-1}}{(i-1)!} \rho^{i-1} + \pi_0 \frac{m^m \rho^m}{m!} \sum_{j=1}^{\infty} (m+j)\rho^j = \\ &= m\rho \sum_{j=0}^{m-1} \pi_j + \pi_0 \frac{m^m \rho^m}{m!} \left(m \sum_{j=1}^{\infty} \rho^j + \sum_{j=1}^{\infty} j\rho^j \right) = \\ &= m\rho(1 - C_m) + \pi_0 \frac{m^m \rho^m}{m!} \left(m \frac{\rho}{1-\rho} + \frac{\rho}{(1-\rho)^2} \right). \end{aligned}$$

Making use of (3.46) we have

$$E[N] = m\rho(1 - C_m) + m\rho C_m + C_m \frac{\rho}{1-\rho}, \tag{3.48}$$

and finally

$$E[N] = m\rho + C_m \frac{\rho}{1-\rho}. \tag{3.49}$$

Using the Little's Result we get

$$E[V] = \frac{1}{\mu} + C_m \frac{1}{m\mu} \frac{1}{1-\rho} = m_X + C_m \frac{m_X}{m(1-\rho)}. \tag{3.50}$$

The Erlang C is related to the Erlang B by

$$C_m(S) = \frac{B_m(S)}{1 - \frac{S}{m}(1 - B_m(S))} \tag{3.51}$$

The Erlang C is greater than the Erlang B as its denominator is less than one. In fact, both provide the probability to find all servers busy, but in the latter case the served traffic served is lesser because blocked users leave. Its behavior with m and a is analogous to that of Erlang B. In particular, with the same traffic per server, S/m, C_m decreases as m increases, and goes to zero when $m \to \infty$ and $S/m < 1$.

The pdf of the waiting time can be found by observing that users in the queue wait, for a new free server, a negative exponential time with rate $m\mu$, i.e., m busy negative exponential servers at rate μ behave like a single negative exponential server at rate $m\mu$. Therefore, from this point of view, users in the queue wait a time that is the same as with one server with rate $m\mu$:

$$f_W(w|W>0) = m\mu(1-\rho)e^{-m\mu(1-\rho)w} \tag{3.52}$$

$$E[W|W>0] = \frac{1}{m\mu(1-\rho)}. \tag{3.53}$$

The unconditional average waiting time is then

$$E[W] = C_m E[W|W>0] = \frac{C_m}{m}\frac{1}{\mu(1-\rho)}. \tag{3.54}$$

Note that, with the same ρ, C_m/m decreases, from ρ with $m=1$, as m increases, meaning that the average time in the queue $E[W]$ decreases. Again, this show the convenience of pooling resources with a single queue, rather than having many queues.

The average time in the system:

$$E[V]_{M/M/m} = E[W] + \frac{1}{\mu} = \frac{C_m}{m\mu(1-\rho)} + \frac{1}{\mu}, \tag{3.55}$$

which is the same as (3.50).

If we compare the system with m servers and the system with a single server m times faster, (3.55) becomes

$$E[V]_{M/M/1} = E[W] + \frac{1}{m\mu} = \frac{\rho}{m\mu(1-\rho)} + \frac{1}{m\mu}. \tag{3.56}$$

Being $C_m \leq \rho$, time $E[W]$ is smaller in the former case, whereas time $E[V]$ is smaller in the latter, and, as m grows, tends to zero.

Example (3.57)
In a system $M/M/2/\infty$ write the close-form distribution and find the probability that a customer has to wait (Erlang C). Find also the average number in the system and average time spent in the system.

For M/M/2/∞ the distribution is

$$\pi_i = \pi_0 \frac{\lambda}{\mu}\left(\frac{\lambda}{2\mu}\right)^{i-1} = \pi_0 2\rho^i \qquad i \geq 1$$

$$\pi_0 + \sum_{i=1}^{\infty} \pi_0 2\rho^i = 1$$

$$\pi_0 = \frac{1-\rho}{1+\rho}$$

$$C_2 = \sum_{i=2}^{\infty} \pi_0 2\rho^i = 2\frac{1-\rho}{1+\rho}\left(\frac{1}{1-\rho} - 1 - \rho\right) = \frac{2\rho^2}{1+\rho}.$$

Note that

$$\frac{2\rho^2}{1+\rho} < \rho.$$

for $\rho \neq 0, 1$. We have also,

$$E[N] = \sum i\pi_i = 2\frac{1-\rho}{1+\rho}\sum_{i=1}^{\infty} i\rho^i = \frac{2}{1+\rho}\frac{\rho}{1-\rho} = 2\frac{\rho}{1-\rho^2}$$

where we have used the know summation

$$(1-\rho)\sum_{i=1}^{\infty} i\rho^i = \frac{\rho}{1-\rho}.$$

Finally, being $\rho = \lambda/2\mu$, we get

$$E[V] = \frac{E[N]}{\lambda} = \frac{1}{\mu}\frac{1}{1-\rho^2}.$$

Also with queueing we can have markovian arrivals represented by the finite population model, that we considered in Section 3.3.2. The solution for these systems can be done on the same lines as above and in Section 3.3.2, and is left to the reader.

To conclude this section we expressly notice the *insensitivity-to-service-pdf* property does not hold for queueing system.

3.6 Markovian Sharing Systems

In Sharing Systems the resources are still limited, but instead of blocking the resource of capacity C is shared among all users. Here we assume that users need a service which is not of the type on/off, as it happens with telephone channel, where the channel is either available or not. Rather, the user is willing to accept a fraction of its intended service speed, or even a greater service speed. This happens, for example, with Internet data calls, where service is referred to the total amount of bits/bytes transferred and not on the time it takes. With this model, if there are n users, then each of them receives the fraction $1/n$ of the bandwidth. Again, sharing changes the shape of the offered traffic curve.

Arrivals occur according to the Poisson Model at rate λ, while service requires a negative exponential pdf with rate $\mu = 1/m_X$ where the rate of service is referred to the unit service capacity. When i active users are present, they share equally the server capacity C, meaning that each receive the fraction C/i of the server capacity.

3.6.1 Complete Fair Sharing

In this model, no user is neither blocked nor queued, but all of them are receiving service; thus, in the limit for large n the service attained by each users goes to zero. With these assumptions the number of active users $N(t)$ is the Birth and Death Process with parameters

$$\lambda_i = \lambda, \qquad \mu_i = \mu C.$$

In fact, with i users the rate of service completion of each user is $\mu C/i$ (fair share), and the global rate of service completion is i times greater. The solution is

$$\pi_i = \pi_0 \frac{\lambda^i}{(\mu C)^i}, \qquad i \geq 0. \tag{3.58}$$

The server load or traffic (utilization factor) is $\rho = \lambda/(\mu C) = \lambda m_X/C$, and normalizing, we get for $\rho < 1$

$$\pi_i = (1-\rho)\rho^i, \qquad i \geq 0, \tag{3.59}$$

that is the same distribution we have in the M/M/1 queuing system. Therefore we have the same average

$$E[N] = \frac{\rho}{1-\rho},$$

and average time in the system, which now represents the average call duration

$$E[V] = \frac{m_X}{C(1-\rho)}.$$

The above clearly shows that $E[V]$ decreases as C increases, going to infinity as $\rho \to 1$.

For $\rho \geq 1$ the solution does not exists, meaning that the chain is not positive-recurrent, the number and the time in the system increase without bounds, each user getting zero service for an infinite amount of time and never departing.

The results above can be generalized as follows (the proof is omitted):

Property *Insensitivity to the service-time pdf:* (3.60)
The above results hold for service times with general pdf.

3.6.2 Response time

We define *response time* to a customer requesting service $X = x$, the average time in the system conditional to service $X = x$:

$$v(x) = E[V/X = x]$$

By the Total Probability Theorem we must have

$$E[v(x)] = \int v(x)b(x)dx = E[V]$$

The FCFS policy provides $v(x) = E[W] + x$, since the queueing time of a user does not depend on its service request x. Things are different for the Processor Sharing Discipline.

Theorem: (3.61)
The response time for the Processor Sharing Discipline is given by

$$v(x) = \frac{x}{1-\rho} \tag{3.62}$$

Proof
Let Δs be the amount of service provided by the server in a time Δt, so small that the number of users does not change within Δt (in fact, there is a probability proportional to Δt that the number changes). The RV service ΔX_i provided to each user at different Δt depends on the number of users in the system during Δt, and, if x represents the amount of service required, we have

$$x = \sum_{i=1}^{R} \Delta X_i$$

where R is the number of service amounts needed by the user. In the same way we can write

$$v(x) = \sum_{i=1}^{R} \Delta t.$$

Using the averages conditional to x we have

$$x = E[R]E[\Delta X_i]. \tag{3.63}$$

$$v(x) = E[R]\Delta t$$

and substituting $E[R]$ from (3.63) into the latter provides

$$v(x) = \frac{x}{E[\Delta X_i]}\Delta t \tag{3.64}$$

The above clearly shows that the time spent in the system is proportional to x and the average service rate $E[\Delta t/[\Delta X_i]]$, which depends on the average number in the system as:

$$E[\Delta X_i] = \frac{\Delta s}{E[N|N>0]},$$

being Δs the overall service provided by the server in Δt. Substituting the latter into (3.64) provides

$$v(x) = x\frac{\Delta t}{\Delta s}E[N|N>0].$$

The ratio

$$\frac{\Delta s}{\Delta t} = C$$

represents the speed of the server, so far implicitly assumed as $C = 1$. The thesis is proven by substituting $E[N|N > 0]$, which by the results on M/M/1, is $1/(1-\rho)$.♣

This result is very interesting because it shows that *a customer that asks for a small service awaits a small time*. The opposite is true for a client who asks a large service.

By Property 3.60, the proof above can be extended to show the following

Property *Insensitivity to the service-time pdf:* (3.65)
Theorem 3.61 holds true for a system M/G/1 with processor sharing.

3.6.3 Sharing with Blocking

We assume that Blocking occurs when there are M users in the system and no more are admitted. The truncation of the state space to M does not alter the equilibrium distribution and, therefore, the equilibrium distribution is the truncated geometric

$$\pi_i = \frac{1-\rho}{1-\rho^{M+1}}\rho^i, \qquad 0 \le i \le M, \tag{3.66}$$

and the blocking probability

$$B = \frac{1-\rho}{1-\rho^{M+1}}\rho^M \tag{3.67}$$

3.6.4 Limited Sharing

Here we consider the case where at each user no more that C/m capacity is assigned, as long as the number of users is less than m. Otherwise the entire capacity C is shared. This correspond to the Internet model where up to m session with a given rate can be accommodated. When congestion occur, i.e., more than m users are present, the capacity is completely shared.

The Markov Chain of the system is exactly the one we saw in section 3.5.2. Denoting again by μ the service completion rate when a unit capacity is assigned, the number of active users $N(t)$ is the Birth and Death Process with parameters

$$\lambda_i = \lambda, \qquad \mu_i = \begin{cases} (i/m)\mu C, & i \le m, \\ \mu C, & i \ge m. \end{cases}$$

Using the server traffic $\rho = \lambda/(\mu C)$, the solution is exactly the one in (erlangC-0)

$$\pi_i = \begin{cases} \pi_0 \dfrac{(m\rho)^i}{i!} & i \le m \\ \\ \pi_0 \dfrac{1}{m!}\dfrac{(m\rho)^i}{m^{i-m}} & i \ge m. \end{cases}$$

Therefore all results about the saturation probability C_m, average number and time in the system are the same as in Section 3.5.2.

In particular, if we denote m_X the average service time with rate $1/(mC)$ we get

$$E[V] = \frac{m_X}{C} + C_m \frac{m_X}{mC(1-\rho)}.$$

3.6.5 Variations of Markovian Systems

In here we deal with more general systems with Markov arrivals and negative exponential service time. It may happen that the occupancy process $N(t)$ is markovian, or it can be represented by another process which is markovian.

Example *M/M/1 with discouraged arrivals* (3.68)
It is a variation of System M/M/1 where the rate of arrival depends on the state of the system: the customer enters service with probability $1/(i+1)$, i being the number of users present in the system. The rates of the Birth and Death process are

$$\lambda_i = \frac{\lambda}{i+1}$$
$$\mu_i = \mu.$$

The solution is

$$\pi_i = \left(\frac{\lambda}{\mu}\right)^i e^{-\lambda/\mu}\frac{1}{i!}.$$

We see it is a Poisson a distribution with average λ/μ. The load factor is

$$\rho_s = 1 - \pi_0 = 1 - e^{-\lambda/\mu}$$

and the average input rate

$$\lambda_{in} = \rho_s/m_x = \mu(1 - e^{-\lambda/\mu})$$

Example *Unequal servers* (3.69)
Let Consider an $M/M/2$ where servers have different service speeds, respectively equal to $\mu_1 = 1$ e $\mu_2 = 2$.

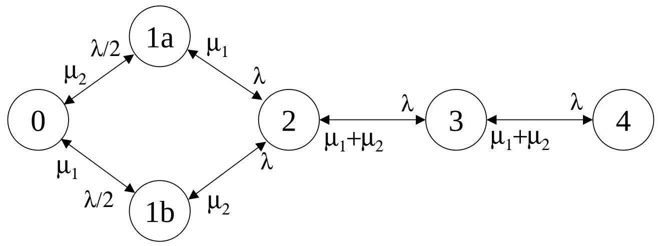
a) Find the queueing probability (i.e., all servers are busy) assuming that, when the system is empty the server is chosen at random. Make evaluations for $\lambda = 2$. Compare then the result with the one attained when both servers are equal with speed $\mu = 3/2$.

b) Find the load of each server.

c) Find the average service time;

d) find the asymptotic rates λ_1^* and μ_1^* for the non-markovian state 1 .

Process $N(t)$ is not markovian since in $N = 1$ the server may be in operation or not, according to the evolution of $N(t)$ in the system. To get a markovian system we must introduce a further state as shown in this state diagram



a) To derive the distribution we observe that from state 2 up, the equations at the cuts are those of a birth and death system. Therefore we have:

$$\pi_i = \pi_2 \left(\frac{\lambda}{\mu_1 + \mu_2} \right)^{i-2}, \qquad i = 2, 3, \ldots$$

and the queueing probability is

$$C = \sum_{i=2}^{\infty} \pi_i = \pi_2 \frac{\mu_1 + \mu_2}{\mu_1 + \mu_2 - \lambda} = 3\pi_2$$

We need π_2. With three cuts in the first part of the diagram we have

$$\begin{cases} \pi_0 = \pi_2 \\ \pi_{1a} = \pi_2/2 \\ \pi_{1b} = \pi_2 \end{cases}$$

Adding $\pi_0 + \pi_{1a} + \pi_{1b} + C = 1$, we get $\pi_2 = 2/11$ e $C = 6/11$.

With equal servers and $\mu_1 = \mu_2 = 3/2$, C is the Erlang C. From this we have $C = 8/15 < 6/11$. Therefore, the equal servers case is better.

b) A way to evaluate the load is to use the Little's result. We evaluate the input/output rate of the first server as

$$\lambda_1 = \mu_1(\pi_{1b} + \sum_{i_2}^{\infty} \pi_i) = 4\pi_2\mu_1 = 8/11$$

Similarly, or subtracting λ_1 from λ, we have $\lambda_2 = 14/11$. The loads are $\rho_1 = 8/11$ and $\rho_2 = 7/11$. An alternative way exploits the meaning of ρ_1 as the probability of the server1 being busy. Therefore

$$\rho_1 = \pi_{1b} + \sum_{i_2}^{\infty} \pi_i = 8/11$$

c) The average service time can be attained by the Little's Result

$$m_x = \frac{\rho_1 + \rho_2}{\lambda} = 15/22$$

or alternatively as

$$\frac{\lambda_1}{\lambda}\frac{1}{\mu_1} + \frac{\lambda_2}{\lambda}\frac{1}{\mu_2} = 15/22$$

d) We see from the figure that the flux toward state 2 is $(\pi_{1a} + \pi_{1b})\lambda = \pi_1\lambda$, and therefore $\lambda_1^* = \lambda = 2$. Similarly, the flux toward state 0 is $\pi_{1a}\mu_2 + \pi_{1b}\mu_1$. This must equal $\pi_1\mu_1^*$. Hence

$$\mu_1^* = \frac{\pi_{1a}\mu_2 + \pi_{1b}\mu_1}{\pi_1} = 4/3.$$

3.7 The departure process

Departure processes represent users leaving a queue system, and are, in a sense, symmetric to the arrival process. We already know that balance conditions imply that the departure frequency is equal to the arrival frequency. In many cases, such a symmetry goes further away. Unfortunately, the interactions introduced by the queue make departures difficult to analyze. This is possible in some cases, for example, the single-server case with Poisson arrivals, since here inter-departure time is rather simple. When the system is busy, inter-departure times coincides with service times X_i, while, when the system is left empty, the inter-departure is the sum of a time Y, the waiting time to the first arrival that make the system work, plus the service time of this arrival. The unconditional inter-departure time can then be evaluated by the Total Probability Theorem.

The evaluation above leads to a close-form pdf in the case of a negative exponential service time. However, a by far more simple proof is possible using reversibility.

3.7.1 Reversible Processes

A process $X(t)$ is said *reversible* if it has the same probabilistic description of process $X(\tau - t)$ for every τ. In other words, a process and its reversed replicas (ie, replicas where the time axis arrow is reversed) are statistically indistinguishable.

Reversibility also implies the stationarity (because of the τ), and, therefore, we now assume $\tau = 0$.

If $X_1, X_2, \ldots, X_n$ are RVs of $X(t)$ taken at increasing time instants $t_1, t_2, \ldots, t_n$, the reversibility requires

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = P(X_1 = x_n, X_2 = x_{n-1}, \ldots, X_n = x_1)$$

for any $t_1, t_2, \ldots, t_n$, and any n.

Markov chains are fully described by the one-step second order distribution and, therefore, the requirement above becomes

$$P(X_{n-1} = j, X_n = k) = P(X_{n-1} = k, X_n = j).$$

or, with the usual notation

$$\pi_j p_{jk} = \pi_k p_{kj}. \tag{3.70}$$

In Chapter 1 we have seen that stationary distributions are such that the total probability fluxes in and out a state must balance. From (3.70) we see that reversibility requires a more stringent

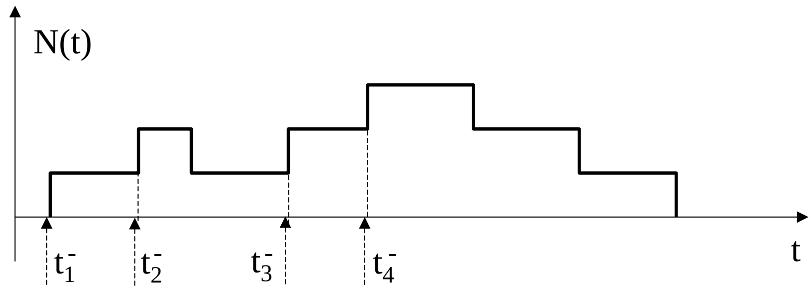


Figure 3.8: *Sample of an occupancy process and arrivals.*

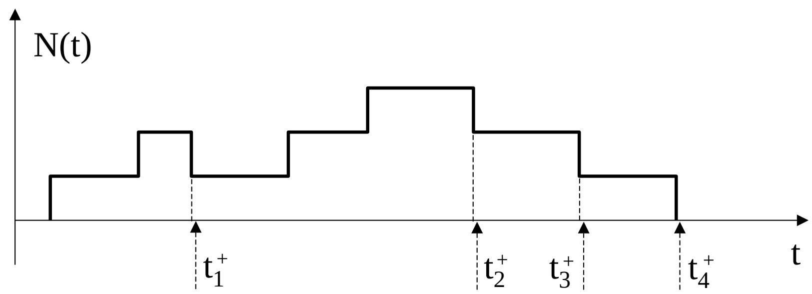


Figure 3.9: *Sample of an occupancy process and departures.*

balance of fluxes, i.e., the balance must exist with fluxes exchanged with any pair of states. This is the case, for example, with processes of the Birth-and-Death type, and in general with processes whose state diagrams are of the *tree* type, ie, diagram where the return path must follow the forward path. For all these processes stationarity also means reversibility, i.e.

Property (3.71)
Stationary MC whose state diagrams are of the tree type are reversible.

However, also processes whose diagram presents closed paths can be reversible:

Theorem: *Kolmogorov* (3.72)
A necessary and sufficient condition for the reversibility of a stationary Markov chain is that, for each state j, the product of the transition probabilities along a closed path that contains j equals the same product along the reverse way. ♣

3.7.2 The departure process of markovian systems

Theorem: *Burke's* (3.73)
In stationary conditions the departures from an $M/M/m/\infty$ system is a Poisson flow with the same characteristics as the arrivals to the same system.

Proof
The proof is very simple and is based on the fact that the occupancy process of an $M/M/m/\infty$

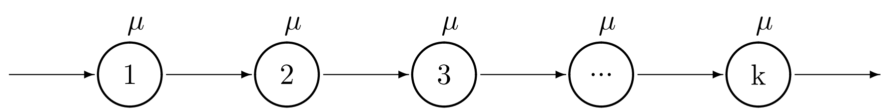


Figure 3.10: *Representation of an Erlang-k service time RV with negative exponential phases.*

system is a Birth-and-Death process and, from what we have seen in the Section 3.7.1, is reversible process. Therefore the process and its reversed present the same probabilistic characteristics.

But in the reversed process arrivals (Figure 3.8) and departures (Figure 3.9) are exchanged, and by the reversibility property are probabilistically the same. Thus also the departure process is Poisson. Obviously, the queue size must be infinite; otherwise the flow into the system is different from arrivals and is not Poisson.

However, if, with a finite queue, users that can not enter the system immediately reach the output flow, and we include such users into the process, then again, the departure process, as the arrival one, is still Poisson.♣

The Burke's theorem, together with union and the splitting properties of a Poisson flows assures that, should users depart as Poisson and entering another markovian system, and so on, all the markovian systems receive Poisson arrivals and, then, can be studied separately as the single systems we saw.

3.8 Toward General Service Time

Let consider an $M/E_k/1$ system, where the pdf of the service time is an Erlang-k, i.e., is the sum of k independent negative exponential random variables of the same rate. The occupation process of this system is not markovian, since Erlang-k is not memoryless. However, the memory associated with this variable is a discrete one. In fact, the service can be seen as a cascade of negative exponential services as shown in Figure 3.10, each phase memoryless (negative exponential), so that the only memory associated with the service time is the current phase of the service (Figure 3.10). What just observed, suggest the enlargement of the state space of an $M/E_k/1$ system as shown in Figure, which is now markovian. We will see the solution of such a system in the next chapter.

More generally a *phase-type* RV can be attained generalizing the representation in Figure 3.10, allowing different paths with different probabilities and negative exponential phases with different rates. Among this class of RV there are sub-classes that can approximate any RV as close as we want. One of such sub-classes is the Cox-type function, schematically represented in Figure 3.12. This means that, potentially, systems with such service time can be reduced to markov systems, although with high complexity.

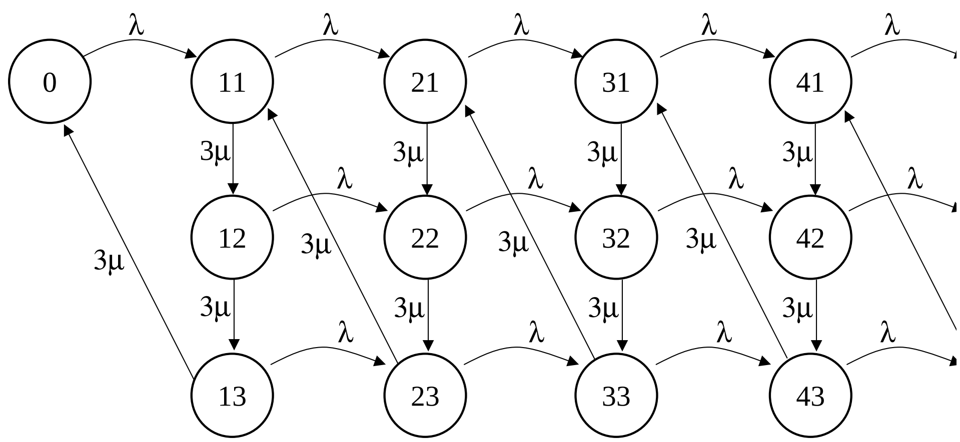


Figure 3.11: *Markovian State-space diagram of an M/E_3/1/ system.*

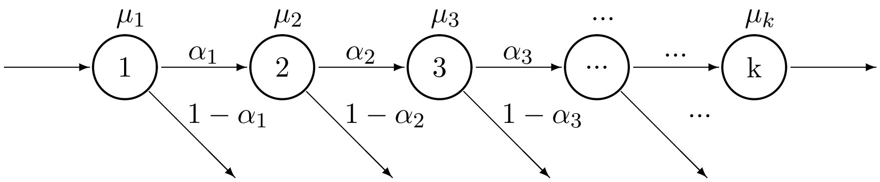


Figure 3.12: *Representation of a Cox-type service time RV.*

3.9 Elements of General Service Time

When the service time is not markovian, then the analysis becomes quite complex, because the occupancy process $N(t)$ itself is no longer markovian. In fact, at any instant t, to determine the end-of-service rate we must know how much time the user currently being served has been in service. This prevents the analysis of many systems. A remarkable case, where the analysis is possible, is the $M/G/1$ system, i.e., a system with one server, general service time, and Poisson arrivals. Here the occupancy distribution and the waiting time pdf's, are known, even if by way of transforms (See Chapter xx), and the results prove that the system is stable if and only if the load factor $\rho = \lambda m_X$ is less than one, in accordance with Theorem 3.38.

Now we show a result, later on generalized to other systems, that can be proved in a simple way.

3.9.1 The Pollaczek-Kinchin formula

Theorem: (3.74)

The average waiting time in an $M/G/1$ system is given by :

$$E[W] = \frac{\rho}{1-\rho} E[Z] \tag{3.75}$$

$$E[Z] = \frac{m_X}{2} + \frac{\sigma_x^2}{2m_X} \tag{3.76}$$

Proof

The average waiting time of a user arriving into the queue presents two components. The first is em remaining service time $E[Z]$, of the user, if any, found in the server. This time corresponds to service that this user must still receive before service completion. The second component, is the time to serve the users found in the queue upon arrival. Denoted by N_c such number, we have

$$E[W] = \rho E[Z] + E[N_c] m_x.$$

Using Little's result we have

$$E[N_c] = \lambda E[W]$$

we get an equation in $E[W]$ that solved provides the (3.75). To find Z we must consider that the outputs of a busy system represents renewal events whose interarrival period coincides with the service times Since a Poisson arrivals behaves as a Random Inspection Point, RV Z V.C. is the one described in the paradox of the renewal events, whose average is equal to the (3.76) . ♣

Here we clearly see that the Renewal Paradox influences the waiting time, which is then worse of what is intuitive. We see an average waiting time that depends on the variance of the service time.

Denoting $c = \sigma_X / m_X$, the (3.75) can be expressed as

$$E[W] = \frac{\rho}{1-\rho} \frac{m_X}{2} (1 + c^2) \tag{3.77}$$

We the have

$$E[V] = m_x + \frac{\rho}{1-\rho}\frac{m_X}{2}(1+c^2) \tag{3.78}$$

$$E[N] = \rho + \frac{\rho^2}{2(1-\rho)}(1+c^2) \tag{3.79}$$

Coefficient c tells when the system behaves better than an $M/M/1$ (case $c < 1$, for example the pdf Erlang-k), or worse ($c > 1$).

Example *M/D/1 System* (3.80)
Being $\sigma_X^2 = 0$ *we have:*

$$E[W] = \frac{\rho}{2(1-\rho)}m_x \tag{3.81}$$

Note that this result is exactly half the case of M/M/1 with equal ρ. This shows that, of the average time spent in the queue by users in an M/M/1 system, half is due to variance in service time, while the other half is due to the variance in the interarrival times.

3.9.2 More user types

Here we introduce differences in users' service time by considering different user types. We refer to an M/G/1 System and use the Pollaczek-Kinchin formula, so that only the mean and the variance of the service time are needed. We exemplify with two user types, but results are easily extended to more.

Users arrive according to Poisson flows, one with rate λ_1 and requires a service time with mean m_1 and variance σ_1^2, whereas the second flow arrives with rate λ_2 and requires a service time with mean m_2 and variance σ_2^2.

The total rate of arrival is $\lambda = \lambda_1 + \lambda_2$. The probability that a user chosen at random is of type 1 is

$$\alpha_1 = \frac{\lambda_1}{\lambda}$$

First and second order moments are

$$m = \frac{\lambda_1}{\lambda}m_1 + \frac{\lambda_2}{\lambda}m_2 \tag{3.82}$$

$$m^{(2)} = \frac{\lambda_1}{\lambda}m_1^{(2)} + \frac{\lambda_2}{\lambda}m_2^{(2)} \tag{3.83}$$

and we get

$$E[Z] = \frac{m^{(2)}}{2m} \tag{3.84}$$

The load factor (traffic) is

$$\rho = \lambda m = \lambda_1 m_1 + \lambda_2 m_2 = \rho_1 + \rho_2$$

The probability that a user in the service is of type 1 is

$$\beta_1 = \frac{\rho_1}{\rho} = \frac{\lambda_1 m_1}{\lambda m}$$

different from α_1 if m_1 is different from m.

The average waiting time in the queue is attained by the Pollaczek-Kinchin formula,

$$E[W] = \frac{\rho E[Z]}{1-\rho}.$$

where $E[Z]$ is provided by (3.84). We must note that the average waiting time does not depend upon the user's type, since the mix of users'types found in the queue is purely random.

An alternative way to derive $E[Z]$ is given by the Total Probability Theorem:

$$E[Z] = \beta_1 E[Z_1] + \beta_2 E[Z_2],$$

where we, again, stress the fact that the weighting probability is not α_i, since it relates to a user found in service. This also shows that the average service time of a user found in the server at time t is

$$m' = \beta_1 m_1 + \beta_2 m_2 = \frac{\lambda_1}{\lambda}\frac{m_1^2}{m} + \frac{\lambda_2}{\lambda}\frac{m_2^2}{m} \tag{3.85}$$

different with respect to (3.82). Again, the (3.82) is the average taken on the population of users, while (3.85) is the average taken on the population of users found in the service in a RIP. These averages are different as long as m_1 is different from m.

The average numbers of users in the queue are

$$E[N_{c1}] = \lambda_1 E[W] \qquad\qquad E[N_{c2}] = \lambda_2 E[W]. \tag{3.86}$$

The average time in the system is

$$E[V] = \frac{\rho E[Z]}{1-\rho} + m$$

whereas

$$E[V_1] = \frac{\rho E[Z]}{1-\rho} + m_1$$

It turns out, obviously,

$$E[V] = \frac{\lambda_1}{\lambda} E[V_1] + \frac{\lambda_2}{\lambda} E[V_2] \tag{3.87}$$

3.10 Problems for solution

P.3.1 A phone makes 8 calls/hour, each lasting an average of 3 minutes.

a) What is the (asymptotic) probability the phone is off?

b) If we have an infinite population of phones that make calls according to Poisson events with the same mean values as above, what is the probability (asymptotic) that no call is on?

P.3.2 An Internet provider has 1000 users, each of which opens an internet session at the average rate of 1.5 every day. The sessions last on average three quarters of an hour, and during each session the user generates packets of 10000 bit length at an average frequency of 300 packets/minute on a channel of 500 kb/s. Find

a) the average number of active sessions;

b) The average number packets being transmitted at the same time by all users;

c) the global average transmission speed.

P.3.3 With reference to the previous exercise, the packets are collected by the provider in a router and transmitted over a channel of 10 Mb/s.

a) find the average traffic on this channel (erlang)

b) the average rate in bit/s?

c) repeat a) and b) with a channel of 20 Mb/s?

P.3.4 (4.17) Two users generate traffic according to a binary MC Markov process $X(t)$ with activation rate λ and service-completion rate μ.

a) Write the stationary distribution of the total traffic when the two users are independent;

b) Write the stationary distribution of the total traffic when, with rate λ the two users become active simultaneously (only when they are both inactive) and cease the activity each with rate μ independently.

c) Find the average traffic in both cases a) and b).

P.3.5 A company originates telephone traffic equal to 2 Erlang, with an average call duration of 5 minutes. It also generates video-conference calls of 0.5 Erlang with an average call duration of 30 minutes. Assuming that the video calls use 4 voice channels simultaneously, find the probability that 5 or more telephone lines are required.

P.3.6 Internet sessions generate traffic flows that are composed of busy periods whose length is exponential distributed with rate μ. Busy periods are separated by idle periods whose length is exponential distributed with rate λ. Find

a) the traffic on the channel, the frequency of activation instants (when the busy period starts), and the average number of activation instants in a period of length τ;

b) assuming stationarity, the probability that the channel is active (busy) at a given time t, and, conditional to this, the probability that the same busy period is still active at $t+\tau$;

c) if the channel is busy at time t the probability that it is still busy (maybe with a different busy period) at time $t+\tau$.

d) the traffic distribution when n equal and independent users are considered.

P.3.7 A transmission channel alternates busy (transmission) and idle periods. Busy periods are independent, negative exponential RVs with rate μ, while idle periods are independent, negative exponential RVs with rate λ. A user wants to transmit on this channel during idle periods, aborting its transmission if the channel becomes busy. Find

a) the probability that the channel is idle at time t, asymptotically.

b) Assuming the channel is sensed idle at time t and that the user starts a transmission of constant duration T in such instant, find the probability that such transmission can be terminated without abortion.

c) repeat point b) when the transmission duration is a negative exponential RV with average T.

P.3.8 (5.12) Referring to an M/M/1 queue with mean service time equal to 1, find the load factor ρ such that the probability of waiting in the queue more that 3 time units is one percent. (Please note that the waiting time in the queue is zero with finite probability).

P.3.9 (5.14) Find the average waiting time, conditional to $W > 0$, in the queue of an M/M/1 system with load $\rho = 0.9$ and average service time equal to 5. Then , find the minimum number of servers must be added (system M/M/m) to reduce the above average to a value no greater than 5.

P.3.10 (5.36) An M/M/1 system with parameters λ and μ has a server that begins to work as soon as the number of users in the system reaches 2 and then continues to work until the system empties. Find

a) the distribution of the number in the system;

b) the average time spent in the system.

P.3.11 (5.1) A queueing system works as follows: if the number of users is less than $N+1$, the server stays idle; the service starts as soon as $N + 1$ users are in the system. During the server activity period, the service time is distributed like a negative exponential with mean $1/\mu$, while new arrived users are rejected until the system is empty, then the process repeats. Assuming the users arrive at the system following a Poisson process with mean rate λ, compute:

a) the asymptotic occupancy distribution;

b) the asymptotic probability that the system is empty;

c) the asymptotic probability that the server is busy;

d) the served traffic.

P.3.12 (5.2) Users arrive at a taxi station according to a Poisson process with mean rate λ. They form a queue if no taxi is available. Taxis arrive at the stop accord to a Poisson process with mean rate $\mu > \lambda$ and, if no customer is waiting, they form a queue. The taxi queue is finite with no more than N positions. Find:

a) the probability a customer waits,

b) the average number of users and taxis in their respective queues,

c) the average waiting time of the users.

P.3.13 (5.5) Users arrive according to a Poisson process with rate λ, and enter a queue without server. The queue empties completely at Poisson instants with rate η.

a) Draw the Markov chain that describes the number $n(t)$ of users in the queue;

b) find the asymptotic distribution;

c) find the average number of users in the queue and its average waiting time.

P.3.14

P.3.15

P.3.16 (5.8) A user enters a M/M/1 queue. Find the probability that he is:

1. the first in a Busy Period;
2. the last in a Busy Period;
3. the first and the last in a Busy Period.

(Hint for b): write the pdf of the sojourn time of the user, and observe that during this time no .....).

P.3.17 (5.16) An M/M/X system is presents a number of servers that changes with arrival and departures of users at the system. If the queue is empty or has 1 user,the system has one server, while when the number in the system is $i \geq 2$ the number of servers is $i-1$. This does not let the queue increase size beyond 1. Determine

a) the stationary distribution of the number of users in the system,

b) the average time spent in the queue and the average time spent in the system.

P.3.18 (5-17) An M/M/1 queue works with this modification: whenever the service ends, the server closes the system (no user can enter the queue) and takes a vacation period distributed like an Erlang-2 with mean value $2/\eta$. After the vacation, the system opens again, users enter and are served. The users arrived during the vacation period are dropped and no longer considered. Determine

a) the diagram of Markov states,

b) the stationary distribution of the number of users in the system;

c) the bound to the arrival rate λ such that the system is positive-recurrent;

d) the average rate of served users and their average time in the system;

e) as in b) when the server's vacation period is a constant equal to T.

P.3.19 (5.31) Users enter a queue with a single server according to a Poisson flow with rate λ. The users' lifetime has a negative exponential pdf with rate μ. At the end of their lifetime, they exit the system, either the queue or the service, and, when in service, the stay in service for their entire lifetime (which means there is no end-of-service, but only end-of-life). Find

a) the stationary distribution of the number of users in the system with the proper equilibrium conditions,

b) the average rate of served users and the related probability of being served,

c) the average rate of users exiting the system when they are on top of the queue and the related probability.

P.3.20 Users arrive according to a Poisson process of rate λ at an M/M/1 system with service rate μ. If a queue builds up, the users entering the queue do not wait for service more than a time Z, random variable, after which they leave the system without getting service.

a) draw the state diagram representing the content of the system when Z is a negative exponential RV with rate ν.
b) find the asymptotic distribution in the special case when $\nu = \mu$, and determine when it exists.
c) find the rate $\lambda' \leq \lambda$ of the users that indeed achieve service (the output frequency of those users is easily evaluated as....);
d) If more than Q users can not be accommodated in the system, find the expression for the blocking probability.

P.3.21 (5.32) Consider a modified M/M/1 queue with infinite queue and user arrivals according to a Poisson process with mean rate λ. At the beginning of each Busy Period, the server decides with probability 0.5 between two possible service rates, μ and 2μ. Find

a) the distribution of the number of users in the system and the related stability condition on parameters λ and μ.

b) the probability that a user finds the server with the slower service rate,

c) the average queueing time of a waiting user (Hint: besides the classical way, one can notice that a user arriving during a Busy Period sees exactly a M/M/1 system until he exits, in both cases).

P.3.22 (5.37) Consider the process $N(t)$, occupation of the sole queue (server excluded) of a M/M/1 system with rates λ and μ. Discuss whether $N(t)$ is Markovian and find the asymptotic distribution.

P.3.23 (5.38) Find the average number of users in the queue of an M/M/2 system with rates λ and μ.

P.3.24 A variation of an M/M/1 queue, with parameters λ and μ, works with this modification: whenever the number in the queue in reaches 2 (i.e. 3 in the system) a server with the same service rate μ is added and starts servicing the users in the queue. This additional server keeps working until one of the servers becomes idle and no one is in the queue, at which time the idle server is withdrawn. Then the procedure repeats.

a) draw the markovian state diagram;
b) find the stationary distribution of the number of users in the system, together with the maximum arrival rate λ allowed for stability;
c) find the traffic served by the additional server while two servers are working;
d) find the average waiting time in the queue and compare with the same in an M/M/2;

P.3.25 Users arrive according to a Poisson process of rate λ at a system where two identical M/M/1 systems, with server rate μ, are present. An arriving user enters the system with the shorter queue, otherwise it selects at random (1/2).

a) draw the bidimensional state diagram representing the content of the two systems when each one allows at maximum three users.

Assume now that, in addition to the behavior above, at the completion of each service, users reach the other system if this decreases the number ahead of him. Note that this procedure leaves the two system either with the same number of users or with a difference of one at most.

b) draw the new bidimensional state diagram representing the content of the two systems when each one allows for an infinite queue;
c) find the distribution of the chain in b);
d) find the probability that the two queues are balanced.

P.3.26 Let Consider an $M/M/2$ where servers have different service speeds, respectively equal to μ_1 and μ_2, $\mu_1 > \mu_2$. Users always choose the faster server when possible, i.e., when both servers are idle upon arrival and when the faster server becomes idle, in which case a possible user in the slower server immediately switch to faster one. Find

a) the asymptotic occupancy distribution;
b) the load of each server;
c) the average service time.

P.3.27 Jobs are served by two processors. When only one job is present both processors serve this one job, thus doubling the processing power. When two jobs are present, immediately each one gets one processor. When more the two jobs are present, only two jobs at a time are served, one job per processor, and the others wait in the queue. Assuming that arrivals are Poisson at rate λ, and their service time is negative exponential with rate μ when served by a single processor, doubled with two processors, find

a) the asymptotic distribution of the jobs in the system;
b) the average waiting time in the queue and in the service respectively and compare with the same figures of an M/M/2 with parameters λ, μ;
c) the asymptotic distribution when both processors are given to the same job only when the job is alone in the system when starting service. In addition neither of the two processors assigned in this way are released before the service ends (the occupation process is not markovian but can be made markovian by....).

P.3.28 A queue has only one server that, when entered by a user, provides service in two cascaded phases, a and b, both of them lasting for independent periods of time, each of them with negative exponential pdf with rate μ. When the server ends the first phase a it immediately begins the second phase b and, when b finishes, the user leaves (this means that the global service time $a+b$ is the sum of two exponential variables, and is not markovian), and a new user, if any in the queue, can enter the server again. Assuming Poisson arrivals, the number in the system (server + queue) n is not markovian. However, we can suitably "enlarge" the state variable n to a new one that takes into account the Phase of the service and that is markovian.

a) draw the markovian state diagram with the new state variable;
b) assuming that a maximum of two users can be in the system (one in the server and one in the queue), find the distribution π_i $i = 0, 1, 2$ of the users in the queue.

P.3.29 Let consider an M/M/m/m system, m even, ie Poisson arrivals at rate λ, m markovian servers at rate μ, no queue (for example, each server is a transmitter of a given speed).
a) Write the distribution of the number in the system.
b) Assume now that each arrival seizes two servers at a time and, when finished, with the same rate μ as before, releases both servers. Again, write the distribution of the number in the system.
c) Assume now that at rate λ_A users of the first type arrive, while at rate λ_B users of the second type (that size two servers) arrive,
c_1) draw the state diagram for the joint number of users $(n_A,\ n_B)$ in the system;
c_2) verify whether the solution is the product form of the distributions in A and B.

P.3.30 Two types of users arrive according to a Poisson process of rate λ_1 (type 1) and λ_2 (type 2) respectively. The type 1 users require one markovian server at a time, which is released with rate μ_1; type 2 users require 2 markovian servers at a time, which are taken together and released together with rate μ_2.

a) Write the traffic generated by the two types of users (i.e., the average number of servers taken if they are always available)
b) If only two servers are available, users that can not match their requirements are blocked and cleared. Draw the Markov chain that represents the number of busy servers;
c) find the blocking probability of the two types of users.

P.3.31 Cars arrive with a Poisson flow of rate λ. Each car carries two users that, upon arrival, reach two different queueing systems a and b, with one markovian server each, independent and of the same rate μ.

a) write the occupancy distribution of queue a;
b) draw the two-dimension markov chain that represent the joint occupancy of the two queues;
c) assuming that the two systems a and b are blocking systems, i.e., no queue and blocked users are cleared, find the joint distribution of users in the two systems;
d) in case c) verify whether the occupancies of the two systems are statistically independent.

P.3.32 (6.3) Two user flows arrive at a single-server queue according to a Poisson process of rates λ_1 and λ_2. They require a negative exponential service times with two different rates, μ_1 μ_2 respectively. Find

a) the average waiting time in the queue for each type of user;
b) the average waiting time for each type of user, and for the average user using a Processor Sharing discipline;

P.3.33 (6.5) Two Poisson user flows arrive with rates λ_1 and λ_2 at a single-server queue. They have constant service times, D and $2D$ respectively. Find

a) the average waiting time in the queue for each type of users,

P.3.34 An $M/M/1$ system, with Poisson arrival at rate λ and service rate μ, can accept arrivals up to $N+1$ and then blocking all arrivals until it reaches state $N-1$, after which arrivals are accepted again up to $N+1$,

a) draw the markovian state diagram of the system;

b) evaluate the blocking probability and the served traffic.

P.3.35 Referring to an M/M/2 system,

a) write the expression for the average time spent in the system;

Now, the two servers are devoted to a single user in the following way. The served user receives in parallel two different and independent services X_1 and X_2, both RV negative exponentially distributed at rate μ. The user leaves the service (meaning that the next in queue enter the service room) when either of the parallel services finishes (i.e., the user leaves when the shorter service ends).

b) Draw the state diagram of the MC number of users in the system and write its asymptotic distribution;

c) Draw the state diagram of the MC used to represent the different case in which the user leaves service when the longer of the parallel services finishes (i.e., both parallel service must be finished).

P.3.36 An M/M/1 system (λ, μ) is modified in such a way that the server changes its service rate (speed), according to the change in the number i in the system, with law $\mu_i = (i+1)\mu, i \geq 1$. Find

a) the asymptotic distribution of the number in the system together with the maximum λ allowed by stability;

b) the server load factor and the average service time;

c) the average waiting time in the queue.

P.3.37 An access point shares the bandwidth in complete fairness among users. If bandwidth requests occur according to a Poisson flow of rate λ, and if service time is negative exponential with rate μ when a user is using the entire bandwidth, find

a) the probability that at a stationary time instant only one user is using the entire bandwidth;

b) the probability that, upon arrival, a user can use the entire bandwidth;

c) the probability that a user is using the entire bandwidth from the beginning to the end of service (hint: he must find the system empty upon arrival, and no one is allowed to arrive during service)

d) the average time in the system, i.e, from the beginning to the end of service.

P.3.38 An M/M/1 system with parameters λ and μ is changed in the following way. Each arrival, at rate λ, is composed by two users that enter the system together only when the content of the system is an even number, zero comprised; otherwise the two users leave.

a) Draw the state diagram of the number of users in the system;

b) find the asymptotic distribution and the condition of its existence;

c) find the rate of arrivals and the rate of served users.

P.3.39 The manager of an M/M/1 systems with a server speed equal to one wants to change the system into an M/M/2 system with equal servers of speed x. Determine x in the different cases in such a way that:

a) the average number in the system must be the same as in M/M/1;

b) the average time in the queue must be the same as in M/M/1.

P.3.40 The M/M/1 system (λ, μ) is modified in the following way. At the arrival of a new user the service rate is set to 2μ, while at the end of each service at rate 2μ the service rate is changed to μ and all subsequent users are served with this rate until a new user arrives, at which time the service rate is set, again to 2μ, and the procedure repeats.

a) Draw the Markov chain that describes the number of users in the system ;

b) find the asymptotic distribution, the conditions for its existence, and the average number in the sistem;

c) find the the average number in the service (traffic) and the average service time.

P.3.41 An M/M/1 system (λ, μ) is modified in the following way. The server only works when an even number of users is present in the system, and remains idle when an odd number of users is present. The server services two users at a time, whose service completes at the same time with rate μ, so that the two served users leave the system at the same time (obviously, only when an even number of users is present).

a) Draw the Markov chain that describes the number of users in the system ;

b) find the asymptotic distribution, and the conditions for its existence;

c) find the average time spent in the queue, and the average time spent in the server.

P.3.42 An M/M/1 system (λ, μ) is modified in the following way. Power failures occur in a memoryless way with rate η. Upon the failure the server stop servicing, the user in the server leaves, the others wait in the queue, and newly arriving users can not enter and leave. Normal operation is restored when power returns. This happens after a negative exponential period of rate ν. Note that power failures/returns can occur even when the system is empty.

a) Draw the Markov chain that describes the number of users in the system ;

b) find the asymptotic distribution, and the conditions for its existence;

c) find the average time spent in the queue (consider that not all arriving users enters the queue);

d) find the probability that an user enters the queue and the probability that it leaves upon failure.

P.3.43 Users arrive to a queue with a single server of rate μ, according to a Poisson flow of rate λ_1. The users only enter the system when they find inside an even number of customers, otherwise leave without service.

a) Find the asymptotic distribution, the served traffic, and the probability that a user enter the system;

A second user flow is added where, at Poisson time instants of rate λ_2, two arrivals occur at the same instant. Again, the users only enter the system when they find inside an even number of customers, otherwise leave without service.

b) Draw the Markov chain that describes the number of users in the system and find the asymptotic distribution with the conditions for its existence;

c) find the offered traffic, the served traffic and the probability that a user cannot enter the system.

P.3.44 Two Poisson flows of users, A and B, of rates λ each, arrive to a markovian queue system with a single server. When the system is empty and an user arrives, the system blocks and clears any user of the opposite type until the system becomes empty again, when the procedure is repeated. In this way, each busy period is composed by users of one type only. The service time of the two user types has rate μ_A and μ_B respectively.

a) Find the asymptotic distribution of the two types of users, jointly and type by type, its average number and the stability conditions;

b) find the offered traffic, the served traffic and the probability that a user is blocked, together and type by type.

3.1 Problems' solutions- Chapter 3 -

P.3.1 A phone makes 8 calls/hour, each lasting an average of 3 minutes.

a) What is the (asymptotic) probability the phone is off?

b) If we have an infinite population of phones that make calls according to Poisson events with the same mean values as above, what is the probability (asymptotic) that no call is on?

Solution a) The probability that the phone is on coincides with the traffic, that is $s = \lambda m_x = 8/60 \times 3 = 0.4$. Therefore, the probability that the phone is off is 0.6.
b) The traffic presents a Poisson distribution; then

$$P(N = 0) = e^{-0.4} = 0.6703$$

P.3.2 An Internet provider has 1000 users, each of which opens an internet session at the average rate of 1.5 every day. The sessions last on average three quarters of an hour, and during each session the user generates packets of 10000 bit length at an average frequency of 300 packets/minute on a channel of 500 kb/s. Find

a) the average number of active sessions;

b) The average number packets being transmitted at the same time by all users;

c) the global average transmission speed.

Solution
a) the average traffic (sessions) per user is: $1.5 \times 3/4/24 = 0.0469$ erlang. the global average traffic (sessions) is 46.9
b) Here the traffic refers to the number of packets. Each packet is active (transmission) for $10000/500 = 20$ ms, the traffic of a single source is $s = 300 \times 0.02/60 = 0.1$, and collectively $S = 46.9 \times 0.1 = 4.69$ packets.
c) $4.69 \times 500 = 2345$ kb/s.

P.3.3 With reference to the previous exercise, the packets are collected by the provider in a router and transmitted over a channel of 10 Mb/s.

a) find the average traffic on this channel (erlang)

b) the average rate in bit/s?

c) repeat a) and b) with a channel of 20 Mb/s?

Solution
a) The 46.9 packets arriving are transmitted on a channel 20 times faster; then the traffic is 20 times lower, ie 0.2345 erlang.
b) the average speed is always 2345 kb/s.
c) The average speed is the same but traffic halves (0.117).

P.3.4 (4.17) Two sources generate traffic according to a binary MC Markov process $X(t)$ with activation rate λ and service-completion rate μ.

a) Write the stationary distribution of the total traffic when the two sources are independent;

b) Write the stationary distribution of the total traffic when, with rate λ the two sources become active simultaneously (only when they are both inactive) and cease the activity each with rate μ independently.

c) Find the average traffic in both cases a) and b).

Solution.
a) If we consider the joint activities of the sources we have four states (00) (10)(01(11). A simpler way is to exploit independence of the sources. Each source is active with probability

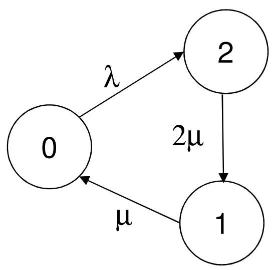
$$\pi_1 = \frac{\lambda}{\lambda + \mu}$$

The traffic has the binomial distribution

$$\pi'_0 = \frac{\mu^2}{(\lambda+\mu)^2} \qquad \pi'_1 = \frac{2\lambda\mu}{(\lambda+\mu)^2} \qquad \pi'_2 = \frac{\lambda^2}{(\lambda+\mu)^2}$$

Alternatively, we can consider the chain of the joint activity of the sources, with three states $0, 1, 2$, representing the active sources. Transition rates are readily derived with the solution, equal as above.

b) Again, if we consider the joint activities of the sources we have the following state diagram. (11). Transition rates are readily found together with the distribution.



and we get

$$\pi''_0 = \frac{2\mu}{3\lambda + 2\mu} \qquad \pi''_1 = \frac{2\lambda}{3\lambda + 2\mu} \qquad \pi''_3 = \frac{\lambda}{3\lambda + 2\mu}$$

c)the average traffic is

$$m = \frac{2\lambda}{\lambda + \mu} \qquad m' = \frac{4\lambda}{3\lambda + 2\mu}$$

P.3.5 A company originates telephone traffic equal to 2 Erlang, with an average call duration of 5 minutes. It also generates video-conference calls of 0.5 Erlang with an average call duration of 30 minutes. Assuming that the video calls use 4 voice channels simultaneously, find the probability that 5 or more telephone lines are required.

Solution.
We need an explicit calculation of the two types of traffic. The event sought splits into three disjoint cases:

a) 0 vide0-calls and 5 or more phone-calls;
b) 1 vide0-call and 1 or more phone-calls;
c) at least 2 vide0-calls.

Taking the infinite population model for the two call types, we have

$$a) \quad e^{-0.5} \times (1 - \sum_{k=0}^{4} \frac{2^k}{k!} e^{-2}) = 0.03193$$

$$b) \quad 0.5e^{-0.5} \times (1 - e^{-2}) = 0.26222$$

$$c) \quad 1 - e^{-0.5} - 0.5e^{-0.5} = 0.09020$$

The sought probability is the sum of the three above.♣

P.3.6 Internet sessions generate traffic flows that are composed of busy periods whose length is exponential distributed with rate μ. Busy periods are separated by idle periods whose length is exponential distributed with rate λ. Find

a) the traffic on the channel, the frequency of activation instants (when the busy period starts), and the average number of activation instants in a period of length τ;
b) assuming stationarity, the probability that the channel is active (busy) at a given time t, and, conditional to this, the probability that the same busy period is still active at $t + \tau$;
c) if the channel is busy at time t the probability that it is still busy (maybe with a different busy period) at time $t + \tau$.
d) the traffic distribution when n equal and independent sources are considered.

Solution
a) The state of the channel is a binary chain with parameters λ and μ. By the definitions we have

$$S = \frac{\lambda}{\lambda + \mu}$$

$$\lambda_a = \frac{1}{m_X + m_Y} = \frac{\lambda\mu}{\lambda + \mu}$$

$$\lambda_a \tau = \frac{\lambda\mu}{\lambda + \mu}\tau$$

b)

$$P(Z = 1) = S = \frac{\lambda}{\lambda + \mu}$$

$$P(Z(t) = 1, \; t_0 < t \leq t_0 + \tau | Z(t_0) = 1) = P(X \geq \tau) = e^{-\mu\tau},$$

where the latter comes from the fact that the duration of the busy period, X is negative exponential, memoryless.

c)

$$P(Z(t + \tau) = 1 | Z(t) = 1) = \pi_1(t + \tau) = \frac{\lambda}{\lambda + \mu}[1 - e^{-(\lambda+\mu)\tau}] + e^{-(\lambda+\mu)\tau}$$

This comes from the transient solution explained in Example 1.77 of the class notes.

d) Since the component traffics are equal and independent the traffic is binomially distributed.

$$P(Z = k) = \binom{n}{k} \left(\frac{\lambda}{\lambda + \mu}\right)^k \left(\frac{\mu}{\lambda + \mu}\right)^{n-k} \qquad (0 \leq k \leq n)$$

P.3.7 A transmission channel alternates busy (transmission) and idle periods. Busy periods are independent, negative exponential RVs with rate μ, while idle periods are independent, negative exponential RVs with rate λ. A user wants to transmit on this channel during idle periods, aborting its transmission if the channel becomes busy. Find

a) the probability that the channel is idle at time t, asymptotically.

b) Assuming the channel is sensed idle at time t and that the user starts a transmission of constant duration T in such instant, find the probability that such transmission can be terminated without abortion.

c) repeat point b) when the transmission duration is a negative exponential RV with average T.

Solution a)

$$\frac{1/\lambda}{1/\lambda + 1/\mu} = \frac{\mu}{\lambda + \mu}$$

b) The sought probability is the probability that the idle lastst for more than T:

$$P(X > T) = e^{-\lambda t}$$

c)

$$P(X > Y) = \frac{1}{\lambda T + 1}$$

P.3.8 (5.12) Referring to an M/M/1 queue with mean service time equal to 1, find the load factor ρ such that the probability of waiting in the queue more that 3 time units is one percent. (Please note that the waiting time in the queue is zero with finite probability).

Solution. The pdf of the average waiting time in the queue can be written as

$$f_W(x) = (1 - \rho)\delta(x) + \rho\mu(1 - \rho)e^{-\mu(1-\rho)x}$$

and

$$P(W > x) = \rho e^{-\mu(1-\rho)x}.$$

Therefore, the sought ρ is provided by the equation ($\mu = 1; x = 3$):

$$\rho e^{-(1-\rho)3} = 0.01,$$

which, provides $\rho = 0.134$.

P.3.9 (5.14) Find the average waiting time, conditional to $W > 0$, in the queue of an M/M/1 system with load $\rho = 0.9$ and average service time equal to 5. Then , find the minimum number of servers must be added (system M/M/m) to reduce the above average to a value no greater than 5.

Solution - The average waiting time in the queue of a customer waits in an M/M/m is given by

$$E[W/W > 0] = \frac{1}{m(1-\rho)}T$$

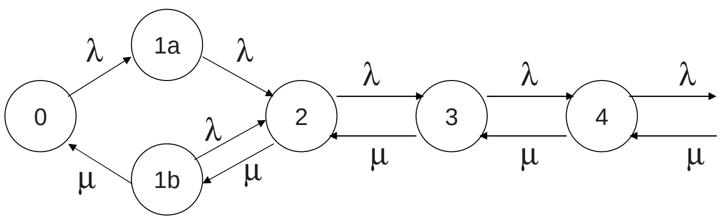
With $\rho = 0.9$ and $m = 1$ we have $E[W/W > 0] = 10T = 50$. To reduce it as above we need at least $m = 10$ servers.

P.3.10 (5.36) An M/M/1 system with parameters λ and μ has a server that begins to work as soon as the number of customers in the system reaches 2 and then continues to work until the system empties. Find

a) the distribution of the number in the system;

b) the average time spent in the system.

Solution - a) The state diagram is shown in the figure.



To get the distribution we observe that from state 2 up, the equations at the cuts are those of a birth and death system. Therefore we have:

$$\pi_i = \pi_2(\lambda/\mu)^{i-2} \qquad i \geq 2$$

We further add

$$\begin{aligned} \pi_2\mu &= (\pi_{1,a} + \pi_{1,b})\lambda \\ \pi_0\lambda &= \pi_{1,b}\mu \\ \pi_0\lambda &= \pi_{1,a}\lambda \end{aligned}$$

Solving, and setting $\pi_1 = \pi_{1,a} + \pi_{1,b}$, we have

$$\pi_i = \pi_0(1 + \lambda/\mu)(\lambda/\mu)^{i-1} \quad i \geq 1$$

$$\pi_0 = (1/2)\ (1 - \lambda/\mu)$$

The distribution exists if and only if $\lambda < \mu$ (since at infinity the system behaves like an M/M/1)

b) The average number is

$$E[N] = \sum i\pi_i = \frac{1}{2}\,\frac{1+\lambda/\mu}{1-\lambda/\mu}$$

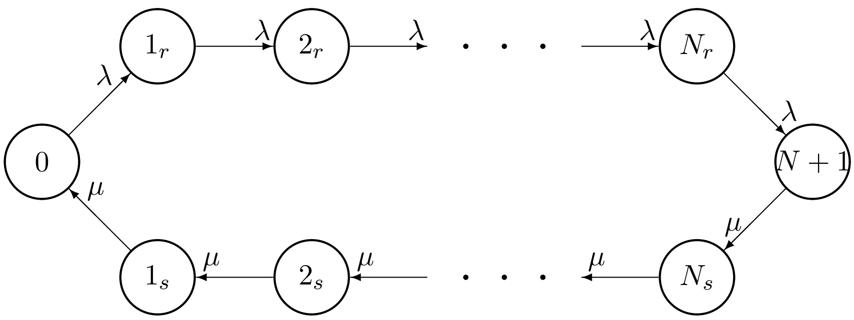
and, from the Little's Result we have

$$E[V] = \frac{1}{2\lambda}\,\frac{1+\lambda/\mu}{1-\lambda/\mu}$$

P.3.11 (5.1) A queueing system works as follows: if the number of users is less than $N+1$, the server stays idle; the service starts as soon as $N+1$ users are in the system. During the server activity period, the service time is distributed like a negative exponential with mean $1/\mu$, while new arrived users are rejected until the system is empty, then the process repeats. Assuming the users arrive at the system following a Poisson process with mean rate λ, compute:

a) the asymptotic occupancy distribution;
b) the asymptotic probability that the system is empty;
c) the asymptotic probability that the server is busy;
d) the served traffic.

Solution The state diagram is shown in the figure



States $i_r, i = 1, \ldots, N$ represent the system in the "filling" phase, where users are accepted without performing any service. States $i_s, i = 1, \ldots, N$ represent the system during the "emptying" phase, where no new customers are accepted and those accumulated in the previous phase are served.

The global balance equations give:

$$\begin{aligned}
\pi_{i_r} &= \pi_0 \\
\pi_{i_s} &= \pi_{N+1} \\
\mu\pi_{N+1} &= \lambda\pi_{N_r} \\
\mu\pi_{1_s} &= \lambda\pi_0
\end{aligned}$$

and the solution is

$$\pi_{i_r} = \pi_0 = \frac{1}{N+1}\,\frac{1}{1+\rho}$$

$$\pi_{i_s} = \pi_{N+1} = \rho\frac{1}{N+1}\,\frac{1}{1+\rho}$$

where $\rho = \lambda/\mu$.

a) The distribution is

$$
\begin{aligned}
P(0) &= \pi_0 = \frac{1}{N+1}\,\frac{1}{1+\rho} \\
P(i) &= \pi_{i_s} + \pi_{i_r} = \frac{1}{N+1} \quad i = 1, 2, \ldots, N \\
P(N+1) &= \pi_{N+1} = \rho\frac{1}{N+1}\,\frac{1}{1+\rho}
\end{aligned}
$$

b) The fraction during which the system is empty is

$$P(0) = \pi_0 = \frac{1}{N+1}\,\frac{1}{1+\rho}$$

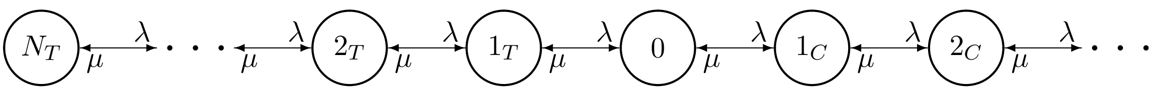
c) The fraction during which the server is busy is

$$\pi_{1_s} + \pi_{2_s} + \ldots + \pi_{N_s} + \pi_{N+1} = \frac{\rho}{1+\rho}$$

P.3.12 (5.2) Customers arrive at a taxi station according to a Poisson process with mean rate λ. They form a queue if no taxi is available. Taxis arrive at the stop accord to a Poisson process with mean rate $\mu > \lambda$ and, if no customer is waiting, they form a queue. The taxi queue is finite with no more than N positions. Find:

a) the probability a customer waits,
b) the average number of customers and taxis in their respective queues,
c) the average waiting time of the customers.

Solution The system state is defined by the number of taxis in the queue or the number of customers in the queue. If we denote by i_T state with i taxi waiting at the station, and by j_C state j where customers are in line await the arrival of the taxi (the state 0 represents the case where there neither taxis nor customers), then the Markov chain has the following transitions diagram:



Except for the state labels, the above diagram coincides with that of an M/M/1 queue and its distribution is

$$
\begin{aligned}
\pi_{i_C} &= (1-\rho)\rho^{i+N} \\
\pi_{j_T} &= (1-\rho)\rho^{N-j}
\end{aligned}
$$

a) A customer waits whenever the system is in states i_C, then the probability for a customer to wait is:

$$\sum_{i=0}^{\infty} \pi_{i_C} = \sum_{i=0}^{\infty} (1-\rho)\rho^{i+N} = \rho^N(1-\rho)\frac{1}{1-\rho} = \rho^N$$

b)

$$\bar{n_C} = \sum_{i=0}^{\infty} i\pi_{i_C} = \sum_{i=0}^{\infty} i(1-\rho)\rho^{i+N} = \rho^{N+1}(1-\rho)\sum_{i=0}^{\infty} i\rho^{i-1} = \frac{\rho^{N+1}}{1-\rho}$$

$$\begin{aligned}\bar{n_T} = \sum_{j=0}^{N} j\pi_{j_T} &= (1-\rho)\rho^N \sum_{j=0}^{N} j\rho^{-j}(1-\rho)\rho^{N-1}\sum_{j=0}^{N} j\left(\frac{1}{\rho}\right)^{j-1} \\ &= (1-\rho)\rho^{N-1}\frac{\rho^{N+1}-1+(N+1)(1-\rho)}{\rho^{N-1}(1-\rho)^2} \\ &= \frac{\rho^{N+1}-(N+1)\rho+N}{1-\rho}\end{aligned}$$

c) Using the Little's result:

$$T = \frac{N_C}{\lambda} = \frac{\rho^{N+1}}{(1-\rho)\lambda} = \frac{\rho^N}{\mu(1-\rho)}$$

P.3.13 (5.5) Users arrive according to a Poisson process with rate λ, and enter a queue without server. The queue empties completely at Poisson instants with rate η.

a) Draw the Markov chain that describes the number $n(t)$ of users in the queue;
b) find the asymptotic distribution;
c) find the average number of users in the queue and its average waiting time.

Solution- a) Chain $n(t)$ presents the following transition rates:

- $q_{i,i+1} = \lambda$
- $q_{i,0} = \eta$.

b)

The balance across node $i+1$ provides

$$\pi_{i+1}(\lambda+\eta) = \pi_i\lambda.$$

This immediately provides

$$\pi_i = \pi_0\left(\frac{\lambda}{\lambda+\eta}\right)^i, \qquad \forall\eta > 0.$$

c) $E[N] = \sum i\pi_i$ can be derived from the distribution. However, a straightforward way is to observe that the waiting time is negative exponential with rate η, and , therefore, by the Little's Result we have $E[N] = \lambda/\eta$.

P.3.14

P.3.15

P.3.16 (5.8) A user enters a M/M/1 queue. Find the probability that:

1. he is the first in a Busy Period;
2. he is the last in a Busy Period;

3. he is the first and the last in a Busy Period.

(Hint for b): write the pdf of the sojourn time of the user, and observe that during this time no .....).

Solution. a) It is clearly $\pi_0 = 1 - \rho$

b) The sojourn time pdf is (see class notes)

$$f_V(x) = \mu(1-\rho)e^{-(\mu(1-\rho)x}$$

If Y is the interarrival time (negative exponential with rate λ we have

$$P(Y > V) = \int P(Y > V/V = x) f_V(x)dx = \int e^{-\lambda x}\mu(1-\rho)e^{-\mu(1-\rho)x}dx =$$

$$= \frac{\mu(1-\rho)}{\lambda + \mu(1-\rho)} = 1 - \rho = \pi_0$$

Is this an accident? (Remember that M/M/1 is reversible..)

c) Similarly

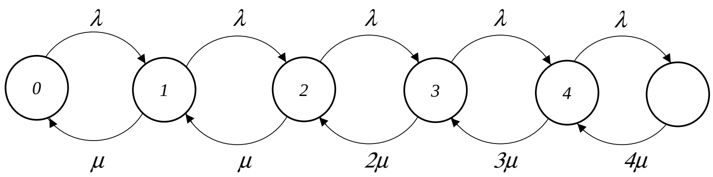
$$P(Y > X) = \int P(Y > X/X = x) f_X(x)dx = \int e^{-\lambda x}\mu e^{-\mu x}dx = \frac{\mu}{\lambda + \mu}$$

P.3.17 (5.16) An M/M/X system is presents a number of servers that changes with arrival and departures of users at the system. If the queue is empty or has 1 user,the system has one server, while when the number in the system is $i \geq 2$ the number of servers is $i - 1$. This does not let the queue increase size beyond 1. Determine

a) the stationary distribution of the number of users in the system,

b) the average time spent in the queue and the average time spent in the system.

Solution - The state diagram is as follows



The distribution is obtained as the solution of a birth and death process, and denoted $\lambda/\mu = \rho$, is :

$$\pi_i = \pi_0 \frac{1}{(i-1)!}\rho^i \qquad i \geq 1$$

with

$$\pi_0 = \frac{1}{1 + \rho\, e^{\rho}}$$

The average number of users is

$$E[N] = \sum_i i\pi_i = \pi_0 \sum_{i=1}^{\infty} \frac{i}{(i-1)!}\rho^i = \pi_0 \rho \frac{d}{d\rho}\left[\rho \frac{d}{d\rho}\left(\sum_{i=1}^{\infty} \rho^i/i!\right)\right] = \rho\ \frac{1+\rho}{e^{-\rho}+\rho}$$

The average time is (Little's result)

$$E[V] = E[N]/\lambda = 1/\mu\ \frac{1+\rho}{e^{-\rho}+\rho}$$

and the average queueing time

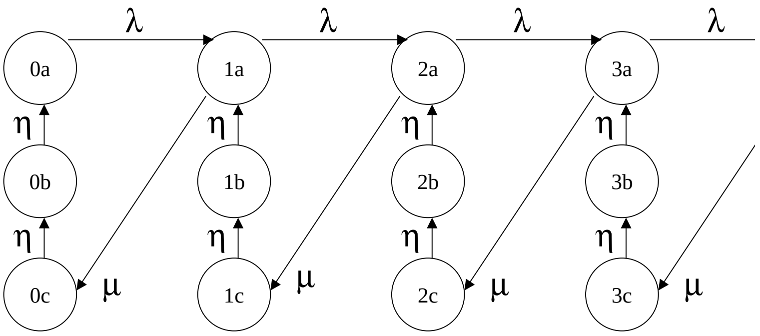
$$E[W] = E[V] - 1/\mu = 1/\mu\ \frac{1-e^{-\rho}}{e^{-\rho}+\rho}$$

Note that $E[W] \leq 1/\mu$, since the queue can never be greater than 1.

P.3.18 (5-17) An M/M/1 queue works with this modification: whenever the service ends, the server closes the system (no user can enter the queue) and takes a vacation period distributed like an Erlang-2 with mean value $2/\eta$. After the vacation, the system opens again, users enter and are served. The users arrived during the vacation period are dropped and no longer considered. Determine

a) the diagram of Markov states,
b) the stationary distribution of the number of users in the system,
c) the bound to the arrival rate λ such that the system is positive recurrent,
d) the rate of served users and their average time in the system.

Solution. a) The state diagram is shown in figure



b) With suitable cuts we have

$$\pi_{ia} = \pi_{0a}(\lambda/\mu)^i \qquad \pi_{ib} = \pi_{ic} = \pi_{i+1a}\mu/\eta = \pi_{ia}\lambda/\eta$$

$$\pi_i = \pi_{ia} + \pi_{ib} + \pi_{ic} = \pi_{0a}(1 + 2\lambda/\eta)(\lambda/\mu)^i$$

$$\pi_{0a} = (1 - \lambda/\mu)/(1 + 2\lambda/\eta)$$

and, finally

$$\pi_i = (1 - \lambda/\mu)(\lambda/\mu)^i$$

It should be noted that the distribution is exactly equal to that of a M/M/1 sysytem; Why? Note also that it does not depend on η.

c) The bound is the one that comes from the convergence of the solution, $\lambda < \mu$.

d) The average number of users is, denoting $\rho = \lambda/\mu$:

$$E[N] = \sum_i i\pi_i = \rho/(1 - \rho)$$

The rate of served users is

$$\lambda_{in} = \sum_i \lambda\pi_{ia} = \lambda/(1 + 2\lambda/\eta)$$

and by the Little's result:

$$E[T] = \rho/(1 - \rho)(1/\lambda)(1 + 2\lambda/\eta)$$

P.3.19 (5.31) Users enter a queue with a single server according to a Poisson flow with rate λ. The users' lifetime has a negative exponential pdf with rate μ. At the end of their lifetime, they exit the system, either the queue or the service, and, when in service, the stay in service for their entire lifetime (which means there is no end-of-service, but only end-of-life). Find

a) the stationary distribution of the number of users in the system with the proper equilibrium conditions,
b) the average rate of served users and the related probability of being served,
c) the average rate of users exiting the system when they are on top of the queue and the related probability.

Solution

a) The distribution in the system is the distribution of the traffic (the lifespan is the activity period) and, therefore is Poisson with average λ/μ. Alternatively, the number of users is a birth and deat process with $\lambda_i = \lambda$ e $\mu_i = i\mu$, whose solution is (Poisson)

$$\pi_k = (\lambda/\mu)^k/k!\ e^{-\lambda/\mu}$$

b) The rate of served users is

$$\sum_{k=1}^{\infty} \mu\pi_k = (1 - \pi_0)\mu = (1 - e^{-\lambda/\mu})\mu$$

and the probability of being served, is the ratio between the rate of served users and the birth rate:

$$\alpha = (\mu/\lambda)(1 - e^{-\lambda/\mu})$$

c) the rate is

$$\sum_{k=2}^{\infty} \mu\pi_k = (1 - \pi_0 - \pi_1)\mu = (1 - e^{-\lambda/\mu} - \lambda/\mu\ e^{-\lambda/\mu})\mu$$

and the related probability is attained by dividing by λ.

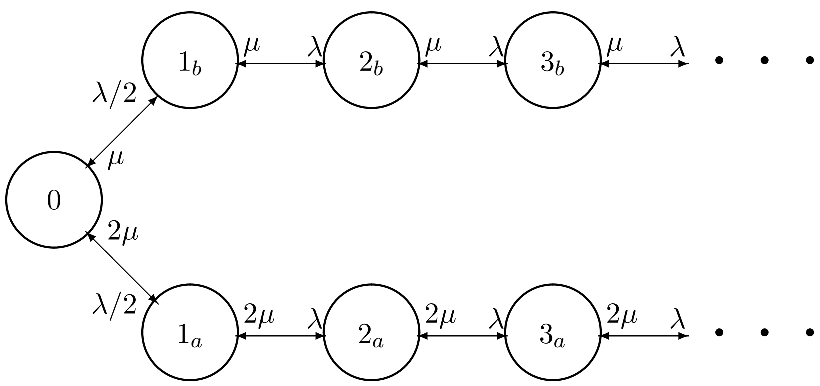
P.3.20 Users arrive according to a Poisson process of rate λ at an M/M/1 system with service rate μ. If a queue builds up, the users entering the queue do not wait for service more than a time Z, random variable, after which they leave the system without getting service.

a) draw the state diagram representing the content of the system when Z is a negative exponential RV with rate ν.
b) find the asymptotic distribution in the special case when $\nu = \mu$, and determine when it exists.
c) find the rate $\lambda' \leq \lambda$ of the users that indeed achieve service (the output frequency of those users is easily evaluated as....);
d) If more than Q users can not be accommodated in the system, find the expression for the blocking probability.

P.3.21 (5.32) Consider a modified M/M/1 queue with infinite queue and user arrivals according to a Poisson process with mean rate λ. At the beginning of each Busy Period, the server decides with probability 0.5 between two possible service rates, μ and 2μ. Find

a) the distribution of the number of users in the system and the related stability condition on parameters λ and μ.
b) the probability that a user finds the server with the slower service rate,
c) the average queueing time of a waiting user (Hint: besides the classical way, one can notice that a user arriving during a Busy Period sees exactly a M/M/1 system until he exits, in both cases).

Solution // a) The state diagram is the following



The distributions of the two branches of the chain depend on π_0, being the solution of the of the birth and death process, and we have, for $i \geq 1$:

$$\pi_{ia} = \pi_0 \frac{1}{2} \left(\frac{\lambda}{2\mu} \right)^i \qquad \pi_{ib} = \pi_0 \frac{1}{2} \left(\frac{\lambda}{\mu} \right)^i$$

$$\pi_0 = \frac{2(2\mu - \lambda)(\mu - \lambda)}{2(2\mu - \lambda)(\mu - \lambda) + \lambda(\mu - \lambda) + \lambda(2\mu - \lambda}$$

b) The sought probability is

$$P_a = \sum_{i=1}^{\infty} \pi_{ia} = \pi_0 \frac{1}{2} \frac{\lambda}{\mu - \lambda}$$

c) the bottleneck is given by the slowest case, so we must have $\lambda < \mu$.

d) We can find the average number in the queue and then use the Little's result. More instructively, we note that a customer sees exactly a system M/M/1 (in either case, it belongs to either busy period). Then we have

$$E[W/W > 0] = P_a E[W_a/W > 0] + P_b E[W_b/W > 0]$$

where

$$E[W_a/W > 0] = \frac{1}{2\mu - \lambda} \qquad\qquad E[W_b/W > 0] = \frac{1}{\mu - \lambda}$$

P.3.22 (5.37) Consider the process $N(t)$, occupancy of the sole queue (server excluded) of a M/M/1 system with rates λ and μ. Discuss whether $N(t)$ is Markovian and find the asymptotic distribution.

Solution $N(t)$ is not Markov because the time spent in state 0 is not memoryless. Denoted by π_i the distribution of the occupancy of the entire system, the distribution sought ν_i is

$$\nu_0 = \pi_0 + \pi_1 \qquad\qquad \nu_i = \pi_{i+1}, \quad i > 0$$

or

$$\nu_0 = 1 - \rho^2 \qquad\qquad \nu_i = (1-\rho)\rho^{i+1}, \quad i > 0$$

with $\rho = \lambda/\mu$.

P.3.23 (5.38) Find the average number of users in the queue of an M/M/2 system with rates λ and μ.

Solution We may proceed by calculating the average number in the system and then subtracting the average number in the services, or directly from the distribution of the queue content.

The distribution of the occupancy of the system is given in the class notes. Otherwise, it can be evaluated by the birth and death process. We have

$$\pi_i = 2\frac{1-\rho}{1+\rho}\rho^i \qquad i \geq 1$$

$$\pi_0 = \frac{1-\rho}{1+\rho}$$

with $\rho = \lambda/(2\mu)$. The average is

$$E[N] = \sum_{i=1}^{\infty} 2i\frac{1-\rho}{1+\rho}\rho^i = 2\frac{1}{1+\rho}\sum_{i=1}^{\infty} i(1-\rho)\rho^i = 2\frac{1}{1+\rho}E[N]_{M/M/1} = \frac{2\rho}{(1-\rho)^2}$$

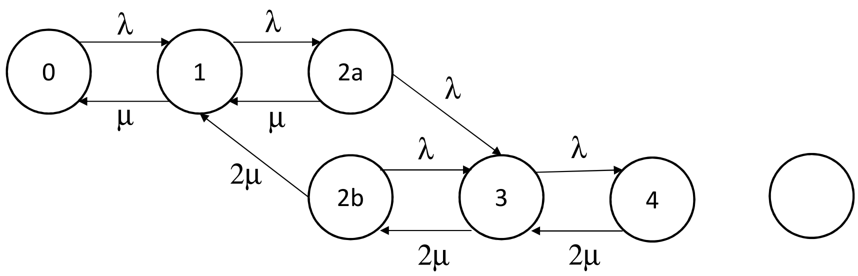
and the average number in the queue is

$$E[N_C] = \frac{2\rho}{(1-\rho)^2} - 2\rho = 2\rho\frac{2\rho-\rho^2}{(1-\rho)^2}$$

P.3.24 A variation of an M/M/1 queue, with parameters λ and μ, works with this modification: whenever the number in the queue in reaches 2 (i.e. 3 in the system) a server with the same service rate μ is added and starts servicing the users in the queue. This additional server keeps working until one of the servers becomes idle and no one is in the queue, at which time the idle server is withdrawn. Then the procedure repeats.

a) draw the markovian state diagram;
b) find the stationary distribution of the number of users in the system, together with the maximum arrival rate λ allowed for stability;
c) find the traffic served by the additional server while two servers are working;
d) find the average waiting time in the queue and compare with the same in an M/M/2;

Solution
a)



b) The distribution can be derived observing that

$$\pi_i = \pi_3 \left(\frac{\lambda}{2\mu}\right)^{i-3}, \qquad i \geq 3.$$

The other π_i are attained by four balance equations and the normalizing equation.

c) Two servers are working when the chain is in states $2b, 3, \ldots$, where the additional server traffic is one. Therefore

$$s = \pi_{2b} + \sum_{i=3}^{\infty} \pi_i$$

d) We must find the average number in the queue

$$E[N_q] = \sum_{i=3}^{\infty} (i-2)\pi_i,$$

$$E[W] = E[N_q]/\lambda$$

For the M/M/2 case refer to the class notes.

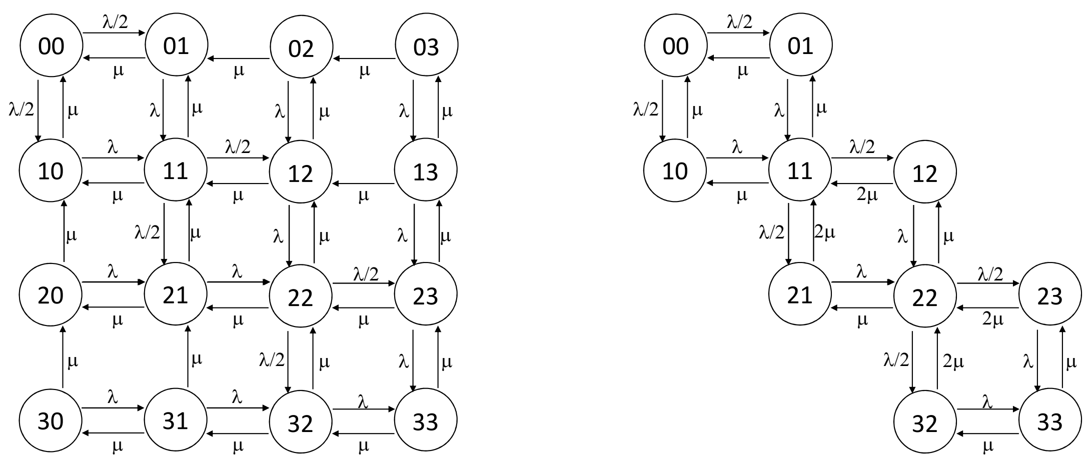
P.3.25 Users arrive according to a Poisson process of rate λ at a system where two identical M/M/1 systems, with server rate μ, are present. An arriving user enters the system with the shorter queue, otherwise it selects at random (1/2).

a) draw the bidimensional state diagram representing the content of the two systems when each one allows at maximum three users.

Assume now that, in addition to the behavior above, at the completion of each service, users reach the other system if this decreases the number ahead of him. Note that this procedure leaves the two system either with the same number of users or with a difference of one at most.

b) draw the new bidimensional state diagram representing the content of the two systems when each one allows for an infinite queue;
c) find the distribution of the chain in b);
d) find the probability that the two queues are balanced.

Solution
The state diagrams corresponding to questions a) and b) are shown below.



The distribution of the rightmost diagram can be easily derived observing that states $(1,2)$ and $(2,1)$ have the same fluxes and ten the same balance equations and, hence, the same probability. The same applies for states $(3,2)$ and $(2,3)$ and so on. This means that there is no need to distinguish them; they can be grouped and the entering flows summed. In another way, from balance equations, the probabilities of those states are naturally grouped two by two. From this we see that states can be re-mapped in this way

$$00 \to 0, \qquad 01, 10 \to 1, \qquad 11 \to 2, \qquad 12, 21 \to 3, \ldots$$

which turns out to be exactly the state diagram of an M/M/2. This result is not surprising since the behavior of the user with the latter modifications is, in fact the same behavior as in M/M/2.

P.3.26 Let Consider an $M/M/2$ where servers have different service speeds, respectively equal to μ_1 and μ_2, $\mu_1 > \mu_2$. Users always choose the faster server when possible, i.e., when both servers are idle upon arrival and when the faster server becomes idle, in which case a possible user in the slower server immediately switch to faster one. Find

a) the asymptotic occupancy distribution;

b) the load of each server;

c) the average service time and the average service time of a client found in the server at a stationary instant.

Solution

The chain is a Birth and Death process with $\lambda_i = \lambda$, $\mu_{1,0} = \mu_1$ e $\mu_{i,i-1} = \mu_1 + \mu_2$, $i > 1$.

a) $\pi_i = \pi_0 \dfrac{\lambda}{\mu_1} \left(\dfrac{\lambda}{\mu_1 + \mu_2} \right)^{i-1}$

b) We must use the Little's result applied to the load factor of each server. Server 1 is busy in states $i = 1, 2, 3...$, whikle server 2 is busy in states $i = 2, 3....$ Therefore

$$\rho_2 = \sum_{i=2}^{\infty} \pi_i = \sum_{i=2}^{\infty} \pi_0 \frac{\lambda^i}{\mu_1(\mu_1 + \mu_2)^{i-1}} = \pi_0 \frac{\lambda^2}{\mu_1(\mu_1 + \mu_2 - \lambda)}$$

$$\rho_1 = \sum_{i=1}^{\infty} \pi_i = \pi_0 \frac{\lambda}{\mu_1} + \pi_0 \frac{\lambda^2}{\mu_1(\mu_1 + \mu_2 - \lambda)} = \pi_0 \frac{\lambda^2 + \lambda(\mu_1 + \mu_2 - \lambda)}{\mu_1(\mu_1 + \mu_2 - \lambda)}$$

By Little's result:

$$\lambda_1 = \rho_1 \mu_1 = \pi_0 \frac{\lambda^2 + \lambda(\mu_1 + \mu_2 - \lambda)}{(\mu_1 + \mu_2 - \lambda)}$$

$$\lambda_2 = \rho_2 \mu_2 = \pi_0 \frac{\mu_2 \lambda^2}{\mu_1(\mu_1 + \mu_2 - \lambda)}$$

We can also derive the service frequencies as

$$\lambda_1 = \sum_{i=1}^{\infty} \mu_1 \pi_i \qquad\qquad \lambda_2 = \sum_{i=2}^{\infty} \mu_2 \pi_i$$

π_0 is derived in the usual way by observing that $\lambda_1 + \lambda_2 = \lambda$.

c) For a randomly chosen user we have

$$m = \frac{\lambda_1}{\lambda} \frac{1}{\mu_1} + \frac{\lambda_2}{\lambda} \frac{1}{\mu_2}$$

At a random point the user in the server is in server 1 with probability ρ_1/ρ, and its service time is $1/\mu_1$; Simuilarly for server 2. Hence,

$$m' = \frac{\rho_1}{\rho} \frac{1}{\mu_1} + \frac{\rho_2}{\rho} \frac{1}{\mu_2}$$

P.3.27 Jobs are served by two processors. When only one job is present both processors serve this one job, thus doubling the processing power. When two jobs are present, immediately each one gets one processor. When more the two jobs are present, only two jobs at a time are served, one job per processor, and the others wait in the queue. Assuming that arrivals are Poisson at rate λ, and their service time is negative exponential with rate μ when served by a single processor, doubled with two processors, find

a) the asymptotic distribution of the jobs in the system;
b) the average waiting time in the queue and in the service respectively and compare with the same figures of an M/M/2 with parameters λ, μ;
c) the asymptotic distribution when both processors are given to the same job only when the job is alone in the system when starting service. In addition neither of the two processors assigned in this way are released before the service ends (the occupation process is not markovian but can be made markovian by....).

Solution
a)

When there is only one job in the system, it is served by two servers and the service rate is, therefore, 2μ. When there are two or more jobs, each one is served by a one server, each one with service rate μ, and, therefore, the state finishing rate is again 2μ. This show that the state diagram is exactly the one of an M/M/1 with descending rate 2μ. The distribution is then

$$\pi_i = (1-\rho)\rho^i, \qquad \rho = \lambda/(2\mu), \qquad i \geq 0.$$

The service time can not be evaluated directly, since the same job can be served at any instant by different number of servers. We use the Little's result by evaluating the average number in the service as

$$E[N_s] = \pi_1 + 2\sum_{i=2}^{\infty}\pi_i = (1-\rho)\left(\rho + 2\sum_{i=2}^{\infty}\rho^i =\right) = (1-\rho)\left(\rho + 2\left(\frac{1}{1-\rho} - 1 - \rho\right)\right) =$$
$$= (1-\rho)\rho + 2 - 2(1-\rho^2) = \rho + \rho^2$$

Notice that for $\rho = 0$ the average is zero, while for $\rho = 1$ the average is two, as it must be. The average service time is then,

$$E[T_s] = \frac{1}{2\mu}(1+\rho),$$

always comprised between $1/(2\mu)$ e $1/\mu$.

The average time in the queue is the difference between the time in the system and the time in the service, where the former is equal to the one in M/M/1, since the average number is the same. Therefore we have

$$E[W] = \frac{1}{2\mu(1-\rho)} - \frac{1}{2\mu}(1+\rho) = \frac{\rho^2}{2\mu(1-\rho)}.$$

In an M/M/2 the average service time is $1/\mu$, while the distribution can be derived directly as shown in Example 3.41. We have

$$\pi_i = \pi_0 2\rho^i \qquad i \geq 1, \qquad \rho = \lambda/(2\mu).$$

$$\pi_0 = \frac{1-\rho}{1+\rho}$$

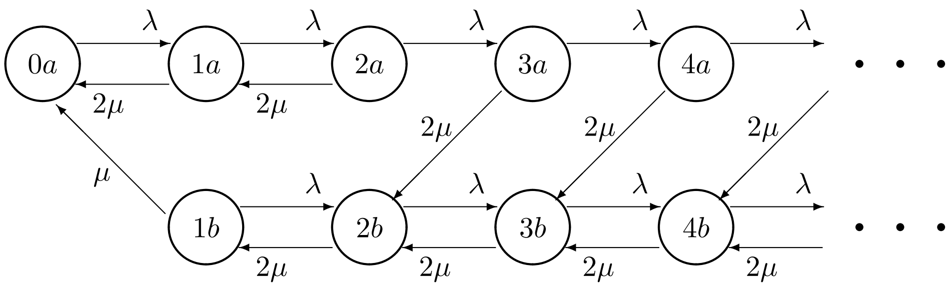
The former shows that the average number in the system is twice that of M/M/1, provided that the correct π_0 is used. Hence

$$E[N] = \frac{\rho}{1-\rho}\frac{2}{1+\rho} = \frac{2\rho}{1-\rho^2}$$

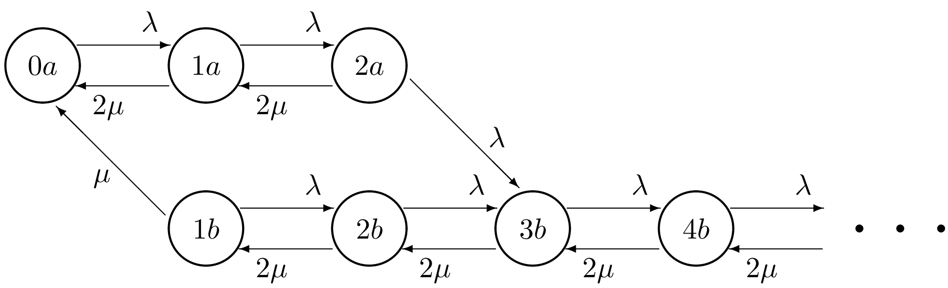
$$E[V] = \frac{1}{\mu(1-\rho^2)}$$

$$E[W] = \frac{\rho^2}{\mu(1-\rho^2)}$$

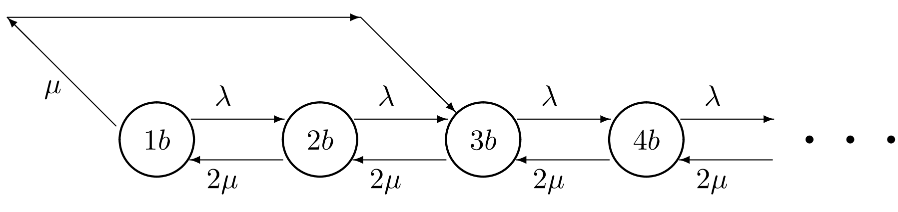
c) the state diagram is



where states a are those where the two servers are dedicated to a single job, and states b otherwise. However, states ia with $i \geq 3$ have the same input and output rates as states ib; therefore we need not distinguish a and b for such states. This yields the following diagram



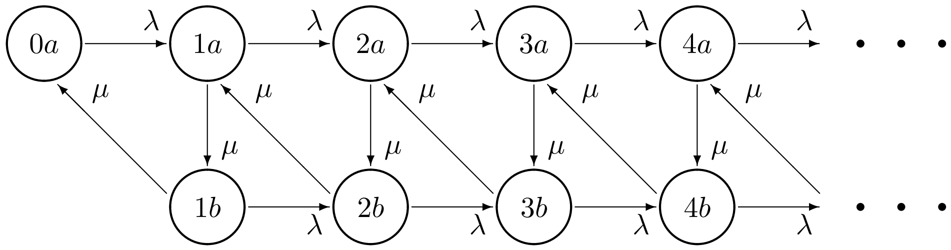
The solution of such a chain can be simplified by solving first the flow of the following chain, which comes from the horizontal cut of the chain above



P.3.28 A queue has only one server that, when entered by a user, provides service in two cascaded phases, a and b, both of them lasting for independent periods of time, each of them with negative exponential pdf with rate μ. When the server ends the first phase a it immediately begins the second phase b and, when b finishes, the user leaves (this means that the global service time $a+b$ is the sum of two exponential variables, and is not markovian), and a new user, if any in the queue, can enter the server again. Assuming Poisson arrivals, the number in the system (server + queue) n is not markovian. However, we can suitably "enlarge" the state variable n to a new one that takes into account the Phase of the service and that is markovian.

a) draw the markovian state diagram with the new state variable;
b) assuming that a maximum of two users can be in the system (one in the server and one in the queue), find the distribution π_i $i = 0, 1, 2$ of the users in the queue.

Solution
a)



P.3.29 Let consider an M/M/m/m system, m even, ie Poisson arrivals at rate λ, m markovian servers at rate μ, no queue (for example, each server is a transmitter of a given speed).
a) Write the distribution of the number in the system.
b) Assume now that each arrival seizes two servers at a time and, when finished, with the same rate μ as before, releases both servers. Again, write the distribution of the number in the system.
c) Assume now that at rate λ_A users of the first type arrive, while at rate λ_B users of the second type (that size two servers) arrive,
c_1) draw the state diagram for the joint number of users $(n_A,\ n_B)$ in the system;
c_2) verify whether the solution is the product form of the distributions in A and B.

Solution
a) It is the case known as Blocking System and the distribution is

$$\pi_i = \frac{A^i/i!}{\sum_{k=0}^{m} A^k/k!} \qquad i = 0, 1, \ldots, m.$$

with $A = \lambda/\mu$.

b) If each user takes two servers, the distribution is the same as before with $m/2$ available resources:

$$\pi_j = \frac{A^j/j!}{\sum_{k=0}^{m/2} A^k/k!} \qquad j = 0, 1, \ldots, m/2.$$

c) If we had $m = \infty$ we would have two non interfering systems (statistically independent), each one with the Poisson distribution, and the joint distribution would be the product. To see if the product form also holds with finite m we should verify if this satisfies the balance equation at a generic node:

$$\pi_{i,j}(\lambda + j\mu + i\mu) = \pi_{i,j+1}(j+1)\mu + \pi_{i+1,j}(i+1)\mu + \pi_{i,j-1}\lambda(1-\alpha) + \pi_{i-1,j}\lambda\alpha$$

which is indeed satisfied by the product of the distributions above. We should also check the balance for border states, and these also satisfy the product form.

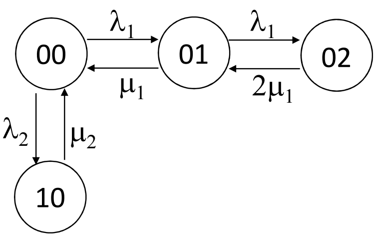
P.3.30 Two types of users arrive according to a Poisson process of rate λ_1 (type 1) and λ_2 (type 2) respectively. The type 1 users require one markovian server at a time, which is released with rate μ_1; type 2 users require 2 markovian servers at a time, which are taken together and released together with rate μ_2.

a) Write the traffic generated by the two types of users (i.e., the average number of servers taken if they are always available)
b) If only two servers are available, users that can not match their requirements are blocked and cleared. Draw the Markov chain that represents the number of busy servers;
c) find the blocking probability of the two types of users.

Solution

a) $\lambda_1/\mu_1 + 2\lambda_2/\mu_2$

b) the state diagram is shown below. States 01 and 02 represent the number of servers taken by traffic of the first type. State 10 represents the state where the two servers are taken by traffic of the second type.



c) The first traffic type is blocked if arrivals find the system in states 02 or 10: $B_1 = \pi_{02} + \pi_{10}$. The second traffic type is blocked if arrivals find the system in states 01, 02 or 10: $B_2 = \pi_{01} + \pi_{02} + \pi_{10} = 1 - \pi_{00}$. We have

$$\pi_{01} = \pi_{00}\frac{\lambda_1}{\mu_1}, \qquad \pi_{02} = \pi_{00}\left(\frac{\lambda_1}{\mu_1}\right)^2, \qquad \pi_{10} = \pi_{00}\frac{\lambda_2}{\mu_2},$$

$$\pi_{00} = \frac{1}{1 + \frac{\lambda_1}{\mu_1} + \left(\frac{\lambda_1}{\mu_1}\right)^2 + \frac{\lambda_2}{\mu_2}}.$$

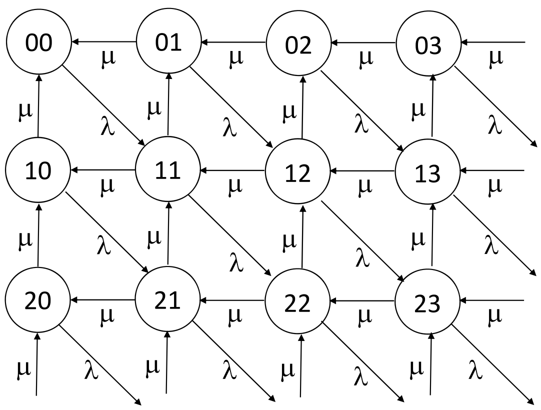
P.3.31 Cars arrive with a Poisson flow of rate λ. Each car carries two users that, upon arrival, reach two different queueing systems a and b, with one markovian server each, independent and of the same rate μ.

a) write the occupancy distribution of queue a;
b) draw the two-dimension markov chain that represent the joint occupancy of the two queues;
c) assuming that the two systems a and b are blocking systems, i.e., no queue and blocked users are cleared, find the joint distribution of users in the two systems;
d) in case c) verify whether the occupancies of the two systems are statistically independent.

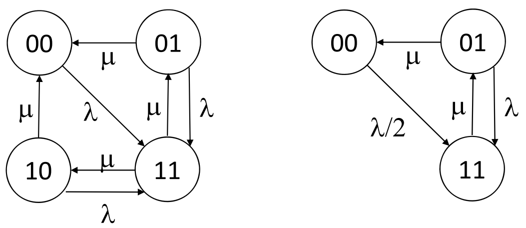
Solution
a) Arrivals at queue a arrive one at a time in coincidence with car arrivals. Therefore, they arrive according to Poisson arrivals with rate λ and queue a) is, in fact, an M/M/1 system with distribution

$$\pi_i = (1-\rho)\rho^i, \qquad \rho = \lambda/\mu$$

b) the state diagram is as follows



c) the state diagram reduces as below left.



Due to the symmetry, the flow balance is the same as in the right part of the figure

$$\pi_{01} = \pi_{10} = \pi_{00}(1/2)\frac{\lambda}{\mu}, \qquad \pi_{11} = \pi_{00}(1/2)\frac{\lambda}{\lambda+\mu}$$

and finally

$$\pi_{00} = \frac{1}{1 + \lambda/\mu + (1/2)(\lambda/\lambda + \mu)}.$$

d) We must find the marginal distributions, that are equal for the two systems. We have

$$\pi'_0 = \pi_{00}\frac{\lambda + 2\mu}{2\mu}, \qquad \pi'_1 = \pi_{00}\frac{\lambda^2 + 2\lambda\mu}{\lambda\mu + \mu^2}.$$

$$\pi''_0 = \pi'_0, \qquad \pi''_1 = \pi'_1$$

We have

$$\pi_{ij} \neq \pi''_i \pi'_j,$$

and the two systems are not statistically independent.

P.3.32 (6.3) Two user flows arrive at a single-server queue according to a Poisson process of rates λ_1 and λ_2. They require a negative exponential service times with two different rates, μ_1 μ_2 respectively. Find

a) the average waiting time in the queue for each type of user;

b) the average time in the system for each type of user, and for the average user using a Processor Sharing discipline;

Solution

a) It is an $M/G/1$ system, where the service pdf is

$$f(x) = \frac{\lambda_1}{\lambda_1 + \lambda_2}\mu_1 e^{-\mu_1 x} + \frac{\lambda_2}{\lambda_1 + \lambda_2}\mu_2 e^{-\mu_2 x}.$$

The average waiting time is

$$E[W] = \frac{\rho E[Z]}{1 - \rho}$$

with

$$E[Z] = \frac{\rho_1}{\rho}E[Z_1] + \frac{\rho_2}{\rho}E[Z_2]$$

$$\rho = \rho_1 + \rho_2 \qquad \rho_1 = \lambda_1/\mu_1 \qquad \rho_2 = \lambda_2/\mu_2$$

$$E[Z_1] = \frac{1}{\mu_1} \qquad E[Z_2] = \frac{1}{\mu_2}.$$

Therefore

$$E[W] = \frac{1}{1 - \rho}\left(\frac{\rho_1}{\mu_1} + \frac{\rho_2}{\mu_2}\right)$$

equal for both types.

b) the class notes say

$$E[W|X = x] = \frac{x}{1 - \rho}$$

and, taking the integration over the two service-time pdf, we get

$$E[W_i] = \frac{1/\mu_i}{1-\rho}, \qquad E[W] = \frac{m_x}{1-\rho}$$

$$m_x = \frac{\lambda_1}{\lambda_1+\lambda_2}\frac{1}{\mu_1} + \frac{\lambda_2}{\lambda_1+\lambda_2}\frac{1}{\mu_2}$$

P.3.33 (6.5) Two Poisson user flows arrive with rates λ_1 and λ_2 at a single-server queue. They have constant service times, D and $2D$ respectively. Find

a) the average waiting time in the queue for each type of users,

Solution
a) The average waiting time is the same for both classes:

$$E[W] = \frac{\rho E[Z]}{1-\rho}$$

To find $E[Z]$ we can refer to the mixed-type service-time pdf is

$$f(x) = \frac{\lambda_1}{\lambda_1+\lambda_2}\delta(x-D) + \frac{\lambda_2}{\lambda_1+\lambda_2}\delta(x-2D),$$

or, more easily, with

$$E[Z] = \frac{\rho_1}{\rho}E[Z_1] + \frac{\rho_2}{\rho}E[Z_2]$$

$$\rho = \rho_1 + \rho_2 \qquad \rho_1 = \lambda_1 D \qquad \rho_2 = 2\lambda_2 D$$

$$E[Z_1] = D/2 \qquad E[Z_2] = D.$$

Therefore

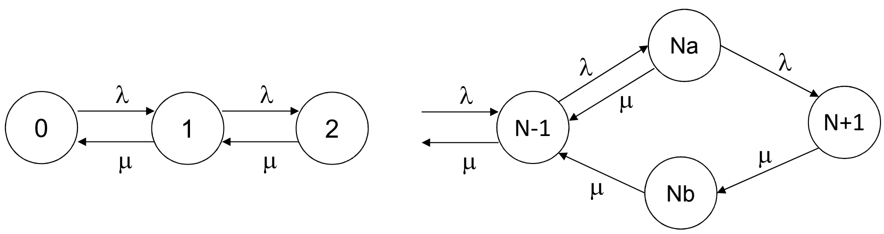
$$E[W] = \frac{\rho_1 + 2\rho_2}{2(1-\rho)}D$$

equal for both types.

P.3.34 An $M/M/1$ system, with Poisson arrival at rate λ and service rate μ, can accept arrivals up to $N+1$ and then blocking all arrivals until it reaches state $N-1$, after which arrivals are accepted again up to $N+1$,

a) draw the markovian state diagram of the system;
b) evaluate the blocking probability and the served traffic.

Solution a) The state diagram is as follows



b) We must find the distribution.

$$\pi_i = \pi_0 \rho^i, \qquad 0 \le i \le N-1,$$

$$\pi_{Na} = \pi_0 \rho^N \frac{1}{1+\rho}, \qquad \pi_{N+1} = \pi_0 \rho^{N+1} \frac{1}{1+\rho}, \qquad \pi_{Nb} = \pi_{N+1}.$$

The blocking probability is

$$P(\text{blocking}) = \pi_{N+1} + \pi_{Nb}.$$

P.3.35 Referring to an M/M/2 system,

a) write the expression for the average time spent in the system;

Now, the two servers are devoted to a single user in the following way. The served user receives in parallel two different and independent services X_1 and X_2, both RV negative exponentially distributed at rate μ. The user leaves the service (meaning that the next in queue enter the service room) when either of the parallel services finishes (i.e., the user leaves when the shorter service ends).

b) Draw the state diagram of the MC number of users in the system and write its asymptotic distribution;

c) Draw the state diagram of the MC used to represent the different case in which the user leaves service when the longer of the parallel services finishes (i.e., both parallel service must be finished).

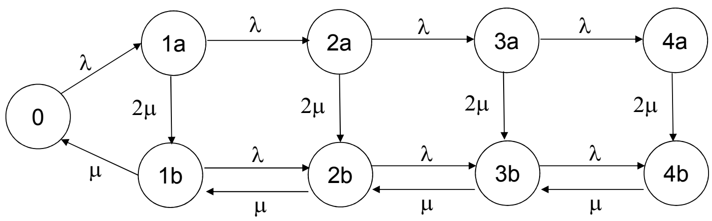
Solution

a) Besides general theory, this very example is shown in the class notes.

$$E[W] = \frac{1}{\mu} \frac{\rho^2}{(1-\rho)^2}.$$

b) The rate of service is practically the same, 2μ, as we have in an M/M/2 system. The difference is that here the rate is still 2μ also when only one user is present. Therefore, the state diagram, and distribution, are the same as with an M/M/1 with service rate equal to 2μ.

c) The state diagram is as shown below, where states a represent the case where the two servers are running, while states b represent the case where only the server with the longer service is running.



P.3.36 An M/M/1 system (λ, μ) is modified in such a way that the server changes its service rate (speed), according to the change in the number i in the system, with law $\mu_i = (i+1)\mu, i \ge 1$. Find

a) the asymptotic distribution of the number in the system together with the maximum λ allowed by stability;
b) the server load factor and the average service time;
c) the average waiting time in the queue.

Solution a) The solution comes from the general solution of the Birth-and-Death process:

$$\pi_i = \pi_0 \frac{(\lambda/\mu)^i}{(i+1)!}, \qquad \pi_0 = \frac{\lambda/\mu}{e^{\lambda/\mu}-1}, \qquad \forall \lambda/\mu.$$

b) The load factor is the probability that the server is busy, which is $1-\pi_0$. The average service time comes from the little's results:

$$E[X] = \frac{1-\pi_0}{\lambda} = \frac{e^{\lambda/\mu}-1-\lambda/\mu}{\lambda(e^{\lambda/\mu}-1)}.$$

c) using $\lambda/\mu = \alpha$ we have

$$E[N] = \sum_{i=1}^{\infty} i\pi_0 \frac{\alpha^i}{(i+1)!} = \pi_0 \alpha \sum_{i=1}^{\infty} i \frac{\alpha^{i-1}}{(i+1)!} = \pi_0 \alpha \sum_{i=1}^{\infty} \frac{d}{d\alpha} \frac{\alpha^i}{(i+1)!} =$$

$$= \pi_0 \alpha \frac{d}{d\alpha}\left(\frac{1}{\alpha}\sum_{i=1}^{\infty} \frac{\alpha^{i+1}}{(i+1)!}\right) = \pi_0 \alpha \frac{d}{d\alpha}\left(\frac{1}{\alpha}(e^\alpha - \alpha - 1)\right) = \pi_0\alpha(e^\alpha/\alpha - e^\alpha/\alpha^2 + 1/\alpha^2) =$$

$$= \frac{1}{e^\alpha - 1}(\alpha e^\alpha - e^\alpha + 1)$$

$$E[W] = E[V] - E[X] = E[N]/\lambda - E[X] = \frac{1}{\lambda}\left(\frac{\alpha e^\alpha - e^\alpha + 1}{e^\alpha - 1} - \frac{e^\alpha - 1 - \alpha}{e^\alpha - 1}\right) =$$

$$= \frac{(e^\alpha + 1)\alpha}{e^\alpha - 1} - 2$$

P.3.37 An access point shares the bandwidth in complete fairness among users. If bandwidth requests occur according to a Poisson flow of rate λ, and if service time is negative exponential with rate μ when a user is using the entire bandwidth, find

a) the probability that at a stationary time instant only one user is using the entire bandwidth;
b) the probability that, upon arrival, a user can use the entire bandwidth;
c) the probability that a user is using the entire bandwidth from the beginning to the end of service (hint: he must find the system empty upon arrival, and no one is allowed to arrive during service)
d) the average time in the system, i.e, from the beginning to the end of service.

Solution
It can be easily seen that, as proven in class notes, the occupancy distribution of a complete-sharing system is exactly the one we get for an M/M/1. Therefore,

a)

$$\pi_1 = (1-\rho)\rho, \qquad \rho = \lambda/\mu.$$

b)

$$\pi_0 = 1 - \rho.$$

c) This question has the same answer as question 3 (or c) in Problem 3.16 of class notes

$$p = \frac{\mu}{\lambda + \mu}.$$

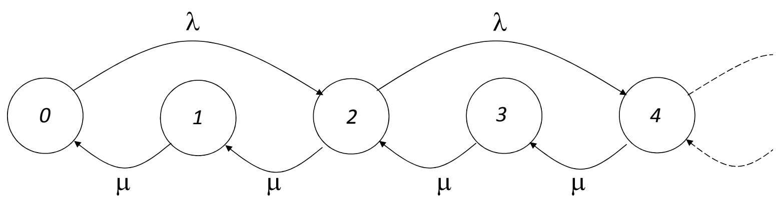
d) This time is exactly the average time spent in the M/M/1 System:

$$E[V] = \frac{1}{\mu}\frac{1}{1-\rho}$$

P.3.38 An M/M/1 system with parameters λ and μ is changed in the following way. Each arrival, at rate λ, is composed by two users that enter the system together only when the content of the system is an even number, zero comprised; otherwise the two users leave.

a) Draw the state diagram of the number of users in the system;
b) find the asymptotic distribution and the condition of its existence;
c) find the rate of arrivals and the rate of served users.

Solution
a) the state diagram is as follows:



b) The distribution can be easily attained as

$$\begin{aligned}
\pi_{2i} &= \pi_0 \left(\frac{\lambda}{\mu}\right)^i, \qquad i = 0, 1, \ldots \\
\pi_{2i+1} &= \pi_{2i+2} \ldots \\
\pi_0 &= \frac{1 - \lambda/\mu}{1 + \lambda/\mu}
\end{aligned} \tag{3.1}$$

and exists for $\lambda < \mu$.
c) The rate of arrivals is clearly 2λ, whereas the rate of served users is

$$\Lambda_s = 2\lambda \sum_i \pi_{2i} = \frac{2\lambda}{1 + \lambda/\mu}.$$

Notice that as $\lambda/\mu \to 1$ we have $\Lambda_s \to \lambda$. What's the explanation?

P.3.39 The manager of an M/M/1 systems with a server speed equal to one wants to change the system into an M/M/2 system with equal servers of speed x. Determine x in the different cases in such a way that:

a) the average number in the system must be the same as in M/M/1;

b) the average time in the queue must be the same as in M/M/1.

Solution a) The M/M/2 system is easily solved to obtain

$$E[N] = \sum i\pi_i = 2\frac{\rho}{1-\rho^2}, \qquad \rho = \frac{\lambda}{2\mu}.$$

whereas in an M/M/1 we have

$$E[N] = \sum i\pi_i = \frac{\rho}{1-\rho}, \qquad \rho = \frac{\lambda}{\mu}.$$

In the M/M/1 case we have $\rho = \lambda$, while in the other case we have $\rho = \lambda/2x$. Then, we must find x that satisfies the following equation:

$$2\frac{\lambda/2x}{1-(\lambda/2x)^2} = \frac{\lambda}{1-\lambda}$$

b) We have seen that in an $M/M/m$ the average time in the queue is the same as in an M/M/1 with server m times as faster. Hence, we must have $x = 1/2$.

P.3.40 The M/M/1 system (λ, μ) is modified in the following way. At the arrival of a new user the service rate is set to 2μ, while at the end of each service at rate 2μ the service rate is changed to μ and all subsequent users are served with this rate until a new user arrives, at which time the service rate is set, again to 2μ, and the procedure repeats.

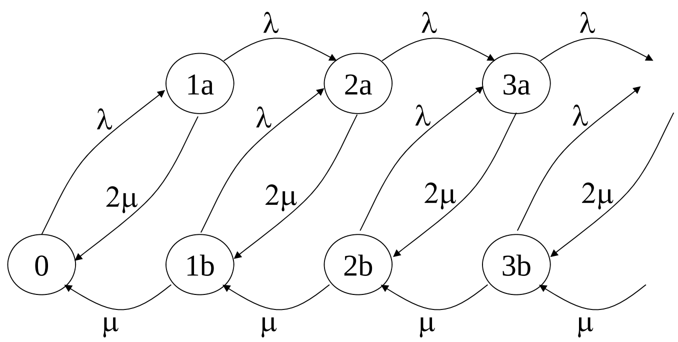
a) Draw the Markov chain that describes the number of users in the system ;

b) find the asymptotic distribution, the conditions for its existence, and the average number in the sistem;

c) find the the average number in the service (traffic) and the average service time.

Solution
a) The state diagram is shown in the figure

Solution a)



b) From a diagonal cut we have

$$\pi_{ia}\lambda = \pi_{ib}\mu, \qquad i \geq 1,$$

and using this into the balance equation to a nodes we have

$$\pi_{i+1,a}(\lambda + 2\mu) = \pi_{i,a}\lambda + \pi_{i,b}\lambda = \pi_{i,a}\lambda + \pi_{i,a}\frac{\lambda^2}{\mu}, \qquad i \geq 1,$$

which provides

$$\pi_{i+1,a} = \pi_{ia}\frac{\lambda(\lambda + \mu)}{\mu(\lambda + 2\mu)}, \qquad i \geq 1.$$

$$\pi_{1,a} = \pi_0 \frac{\lambda}{(\lambda + 2\mu)}.$$

Denoted

$$\eta = \frac{\lambda(\lambda + \mu)}{\mu(\lambda + 2\mu)},$$

we can write

$$\pi_{ia} = \pi_0 \frac{\mu}{\lambda + \mu}\eta^i \quad i \geq 1,$$

$$\pi_{ib} = \pi_0 \frac{\lambda}{\lambda + \mu}\eta^i \quad i \geq 1,$$

and also

$$\pi_i = \pi_{ia} + \pi_{ib} = \pi_0 \eta^i \quad i \geq 0.$$

and finally

$$\pi_0 = (1 - \eta),$$

where the convergence of the solution exists for $\eta < 1$ or $\lambda/\mu < \sqrt{2}$.

The average number in the system is

$$E[N] = \sum_{i=1}^{\infty} i\, \pi_i = \frac{\eta}{1 - \eta}.$$

c) The average number in the service, coincides with the probability the server is busy, $1 - \pi_0 = \eta$, and the average time in service is

$$(1 - \pi_0)/\lambda = \frac{\eta}{\lambda} = \frac{1}{\mu}\frac{(\lambda + \mu)}{(\lambda + 2\mu)},$$

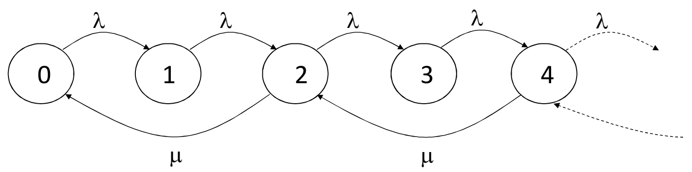
shorter than $1/\mu$.

P.3.41 An M/M/1 system (λ, μ) is modified in the following way. The server only works when an even number of users is present in the system, and remains idle when an odd number of users is present. The server services two users at a time, whose service completes at the same time with rate μ, so that the two served users leave the system at the same time (obviously, only when an even number of users is present).

a) Draw the Markov chain that describes the number of users in the system ;
b) find the asymptotic distribution, and the conditions for its existence;
c) find the average time spent in the queue, and the average time spent in the server.

Solution
a) The state diagram is shown in the figure

Solution a)



b) In odd states the fluxes, in and out, are equal. Then we have

$$\pi_i = \pi_{i-1} \qquad\qquad i = 1, 3, 5, 7, \ldots$$

$$\pi_i = \pi_0 \left(\frac{\lambda}{\mu}\right)^{i/2} \qquad\qquad i = 2, 4, 6, \ldots$$

With the notation $\rho = \lambda/\mu$, which is the offered traffic, by $\sum_i \pi_i$ we get

$$\pi_0 = \frac{1}{2\sum_{i=0}^{\infty} \rho^i} = \frac{1-\rho}{2},$$

which exists only for $\rho < 1$. Note that for $\rho = 0$ we get $\pi_0 = 1/2$, why?

c) The average number and time in the system are

$$E[N] = \pi_0 \sum_{i=0}^{\infty} (2i+2i+1)\rho^i = \pi_0 \left(4\sum_{i=0}^{\infty} i\rho^i + \sum_{i=0}^{\infty} \rho^i\right) = \pi_0 \left(\frac{4\rho}{(1-\rho)^2} + \frac{1}{1-\rho}\right) = \frac{2\rho}{1-\rho} + \frac{1}{2},$$

$$E[V] = \frac{1}{\mu}\frac{2}{1-\rho} + \frac{1}{2\lambda}.$$

The average number and time in the server are

$$E[N_s] = 2\sum_{i=2}^{\infty} \pi_i = 2(1 - \pi_0 - \pi_1) = 2 - 4\pi_0 = 2\rho$$

$$E[T_s] = 2\rho/\lambda = 2/\mu.$$

P.3.42 An M/M/1 system (λ, μ) is modified in the following way. Power failures occur in a memoryless way with rate η. Upon the failure the server stop servicing, the user in the server leaves, the others wait in the queue, and newly arriving users can not enter and leave. Normal operation is restored when power returns. This happens after a negative exponential period of rate ν. Note that power failures/returns can occur even when the system is empty.

a) Draw the Markov chain that describes the number of users in the system ;
b) find the asymptotic distribution, and the conditions for its existence;
c) find the average time spent in the queue (consider that not all arriving users enters the queue);
d) find the probability that an user enters the queue and the probability that it leaves upon failure.

P.3.43 Users arrive to a queue with a single server of rate μ, according to a Poisson flow of rate λ_1. The users only enter the system when they find inside an even number of customers, otherwise leave without service.

a) Find the asymptotic distribution, the served traffic, and the probability that a user enter the system;

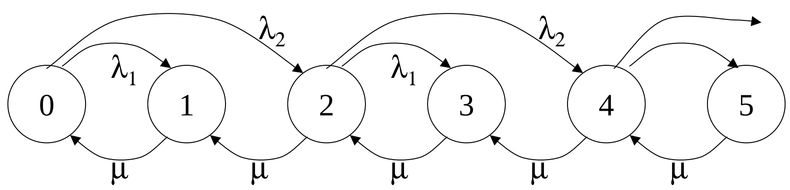
A second user flow is added where, at Poisson time instants of rate λ_2, two arrivals occur at the same instant. Again, the users only enter the system when they find inside an even number of customers, otherwise leave without service.

b) Draw the Markov chain that describes the number of users in the system and find the asymptotic distribution with the conditions for its existence;
c) find the offered traffic, the served traffic and the probability that a user cannot enter the system.

Solution
a) Users can enter only when the system is empty. Therefore, we have a two state chain (λ, μ), whose distribution is

$$\pi_0 = \frac{\mu}{\lambda + \mu}, \qquad \pi_1 = \frac{\lambda}{\lambda + \mu}.$$

The served traffic is π_1, or $\lambda/\mu\pi_0$, and the probability that a user enter the system is π_0.
b) The state diagram is as follows:



The distribution is easily attained taking vertical cuts, and distinguishing between even and odd states. We get

$$\pi_{2i} = \left(\frac{\lambda_2}{\mu}\right)^i \pi_0 \qquad i \geq 0$$

$$\pi_{2i+1} = \frac{\lambda_1 + \lambda_2}{\mu} \left(\frac{\lambda_2}{\mu}\right)^i \pi_0 \qquad i \geq 0$$

$$\pi_0 = \frac{1 - \lambda_2/\mu}{1 + \lambda_1/\mu + \lambda_2/\mu}$$

The convergence condition is $\lambda_2 < \mu$.

c) The offered traffic is $(\lambda_1 + 2\lambda_2)/\mu$. The served traffic is

$$(\lambda_1 + 2\lambda_2)/\mu \sum_i \pi_{2i} = (\lambda_1 + 2\lambda_2)/\mu \; \pi_0 \frac{1}{1 - \lambda_2/\mu} = \frac{(\lambda_1 + 2\lambda_2)/\mu}{1 + \lambda_1/\mu + \lambda_2/\mu}$$

and the probability that a user cannot enter the system is

$$\sum_i \pi_{2i+1} = 1 - \sum_i \pi_{2i} = \frac{\lambda_1/\mu + \lambda_2/\mu}{1 + \lambda_1/\mu + \lambda_2/\mu}$$

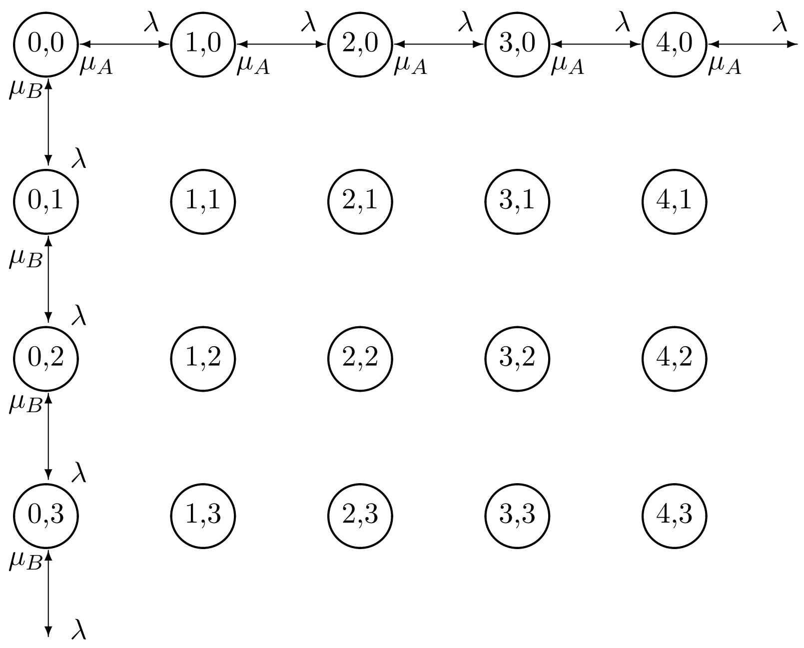
P.3.44 Two Poisson flows of users, A and B, of rates λ each, arrive to a markovian queue system with a single server. When the system is empty and an user arrives, the system blocks and clears any user of the opposite type until the system becomes empty again, when the procedure is repeated. In this way, each busy period is composed by users of one type only. The service time of the two user types has rate μ_A and μ_B respectively.

a) Find the asymptotic distribution of the two types of users, jointly and type by type, its average number and the stability conditions;

b) find the offered traffic, the served traffic and the probability that a user is blocked, together and type by type.

Solution

a) The state diagram is as follows



Denoting $\rho_A = \frac{\lambda}{\mu_A}$ and $\rho_B = \frac{\lambda}{\mu_B}$, the joint distribution presents only the following elements

$$\pi_{i0} = \pi_{00}\rho_A^i, \qquad i \geq 0 \qquad\qquad \pi_{0j} = \pi_{00}\rho_B^j, \quad j \geq 0.$$

Then, π_{00} is obtained from:

$$\sum_{i=0}^{\infty} \pi_{i0} + \sum_{j=1}^{\infty} \pi_{0j} = 1$$

The marginal is

$$\pi_{i0} = \pi_0 \rho_A^i, \qquad i \geq 0, \qquad\qquad \pi_0 = \sum_{j=0}^{\infty} \pi_{0j} = 1 - \rho_A.$$

b) The probability that a type-A user is blocked is

$$P_A = \sum_{j=1}^{\infty} \pi_{0j} = \pi_{00}\frac{\rho_B}{1-\rho_B},$$

the served rate $\lambda(1 - P_A)$, and the served load $\rho_A(1 - P_A)$.

For a generic user we have

$$P = \frac{1}{2}(P_A + P_B),$$

the served rate $\lambda(1 - P_A) + \lambda(1 - P_B)$, and the served load $\rho_A(1 - P_A) + \rho_B(1 - P_B)$.

Chapter 5

Servicing Policies

5.1 Service disciplines in an M/G/1

Service disciplines are policies for selecting users for service. The natural policy is the one known as the *first-come, first-served* (FCFS). Other policies are for example, the LCFS (*last come, first served*), or RO (*random order*). These policies are called *Service Independent*, because the selected user is chosen regardless of the service. Other policies, such as the SJF (*shortest job first*) are based on the specific service requirement x_i of the different user in the queue. Others, again, allow the interruption of the service before completion, and the replacement with another user, and are called *preemptive policies.* A further important classification relates to the work present in the system: if the work can not be reduced other than by service, and created other than by arrivals, the system is said *work-conserving.* The latter is the only class we refer next.

We are interested in knowing to what extent policies are able to change the behavior of the queue and influence waiting times.

If we consider the succession of idle (I) and busy periods (B) in a sample of $N(t)$, it can be seen that

Property (5.1)
none among the cited service disciplines can change the length of any busy period.

In fact, at any time the work present in the system is not changed by the cited disciplines and, therefore the time it takes to service that work is not changed in turn. This also shows that the number of arrivals a in a busy period can not be changed, and proves the property above.

In an M/G/1 system the sequence of idle and busy periods is a regenerative process, and the asymptotic probability to be in a busy period, which is equal to ρ, can also be written as

$$\rho = \frac{E[B]}{E[I] + E[B]} = \frac{E[B]}{1/\lambda + E[B]},$$

where Poisson arrivals assure that $E[I] = 1/\lambda$. Solving for $E[B]$ provides

$$E[B] = \frac{m_X}{1-\rho}. \tag{5.2}$$

Denoted by $E[U]$ the average number of users serviced in a busy period we have $E[B] = m_x E[N_u]$, and using (5.46) we get

$$E[N_u] = \frac{1}{1-\rho} \tag{5.3}$$

If the length of the busy period, and hence the number of served users in the busy period, can not be changed by service disciplines, the average $E[N(t)$ can be changed by some. For example, using SJF users are served more quickly at the beginning of the busy period so that, being the number of users in the busy period the same, the average $E[N(t)$ diminishes, together with $E[V]$ and $E[W]$. This can not happen with service-independent policies, where service time is drawn at the moment the user enters service. Then, we have

Property (5.4)
all policies that are Service Independent and non-preemptive can not alter the distribution of $N(t)$ and therefore, neither can change $E[N]$, $E[V]$ and $E[W]$.

We notice, however, that the service policies above can alter the distributions of V e W, even though the average is not altered. In fact, it is possible to derive the distribution for the LCFS discipline. For example we have

$$\mathrm{VAR}[W]_{FCFS} = (1-\rho)\mathrm{VAR}[W]_{LCFS} - \rho E[W]^2$$

Additional quality measures are represented by the conditional times $E[V/X = x]$ and $E[W/X = x]$, i.e., that are related to the required service amount $X = x$. For the service-independent disciplines we have

$$E[V/X = x] = E[V],$$

since the service requirement x is not taken into account. In some applications this is not considered fair, and, therefore, it is interesting to look for disciplines where $E[V/X = x]$ is more or less proportional to x. This can be achieved using *preemptive* service disciplines where service can be interrupted, the user in service placed in the queue, and a new user is selected for service.

An interesting policy able to achieve this goal is the *time-sharing* policy. With this policy, each customer in the system is placed in service for a small amount of time Δt, after which it reaches the end of the queue and the user at the head of the line is placed into service, and the procedure repeats. In this way a user that requires greater service time than another is forced to take a larger number of service cycles, and therefore, to suffer larger time in the queue and in the system.

A procedure of this type is the one used by the computers' operating systems. If the quantum of service Δt becomes infinitesimal, the procedure coincides with the one known as *processor sharing* (PS), in which the queue disappears and all the n users in the system take a fraction of $1/n$ of the server power.

5.2 Symmetric Queues

The insensitivity properties to the service-time pdf, seen in Section 3.5 for sharing systems, can be extended to a more general family of service disciplines called *Symmetric Queues*. In order to deal with servers with different service speed we must decouple the required service, χ, from the service speed ϕ. We require

a) the service requirement is RV χ with average m_χ;

b) the overall service rate with i users is $\phi(i) > 0$;

c) a fraction $\gamma(l,i), \sum_l \gamma(l,i) = 1$, of the service speed is given to the users u_l in position l in the queue, $1 \leq l \leq i$; when this user leaves service users in positions $l+1, l+2, \ldots, i$, move in positions $l, l+1, \ldots, i-1$;

d) when a user arrives at the queue with i users, it enters the system in position l, $l = 1, \ldots, i+1$ with probability $\delta(l, i+1), \sum_l \delta(l,i) = 1$; users previously in positions $l, l+1, \ldots, i$, move in positions $l+1, l+2, \ldots, i+1$; respectively;

This framework is able to represents many service disciplines. For example

i - FCFS M/G/1 by setting $\phi(i) = \phi$, $\gamma(1,i) = 1$, $\gamma(l,i) = 0, l \neq 1$, $\delta(i,i) = 1$, $\delta(l,i) = 0, l \neq i$; the average service time is $m_X = m_\chi/\phi$.

ii - LCFS M/G/1 with preemption by setting $\phi(i) = \phi$, $\gamma(1,i) = 1$, $\gamma(l,i) = 0, l \neq 1$, $\delta(1,i) = 1$, $\delta(l,i) = 0, l \neq 1$;

iii - Random Order M/G/1 with preemption by setting $\phi(i) = \phi$, $\gamma(1,i) = 1$, $\gamma(l,i) = 0, l \neq 1$, $\delta(l,i) = 1/i, \quad \forall l$;

iv - M/G Processor sharing by setting $\phi(i) = \phi$, $\gamma(1,i) = 1/i, \forall l$;

v - M/G/∞ by setting $\phi(i) = i\phi$ e $\gamma(l,i) = 1/i$; each user has its own server at rate ϕ.

vi - M/G/m/m by setting

$i \leq m$ - $\phi(i) = i\phi$ e $\gamma(l,i) = 1/i$; each user has its own server at rate ϕ.

$i > m$ - $\phi(i) = \infty$, $\gamma(l,i) = 0, l < m$, $\gamma(l,i) = \frac{1}{i-m}, m < l \leq i$; this deliver an infinite service rate to users that upon arrival find m users in the system; this causes these users to immediately leave the system. This account for the fact that such users are, in the real system blocked and cleared.

Symmetric queues are attained by the preceding framework by setting $\delta(l,i) = \gamma(l,i)$ (hence the term symmetric). Unfortunately, this constraint can not model such important system as FCFS, whereas they can model for example PS and LCFS with preemption. In fact, in symmetric queues the service discipline is such that all users in the system have already received some amount of service. In fact, any new arrival changes the service speed of the others, and then is an preemptive one.

For symmetric queues the following results hold (the proof is omitted):

Theorem: (5.5)
the occupancy distribution is given by

$$\pi_i = \pi_0 \frac{(\lambda m_\chi)^i}{\prod_{l=1}^{i} \phi(l)} \tag{5.6}$$

We note that (5.6) does not depend on $\gamma(l,i)$. Furthermore, distribution (5.6) represents the solution of the Birth and Death process (1.154).

With one constant-speed server (5.6) becomes

$$\pi_i = (1-\rho)\rho^i \quad i \geq 0 \tag{5.7}$$

with $\rho = \lambda m_\chi/\phi$, exactly as in M/M/1. Then, we have

Property (5.8)
the occupancy distribution and the average waiting times of an M/G/1 with symmetric queues is the same as with FCFS.

With m servers and no queue the (5.6) becomes

$$\pi_i = \pi_0 \frac{(m\rho)^i}{n!} \quad i \leq m, \tag{5.9}$$

again equal to the markovian case. Here the symmetry constraint has no effect and the solution is exactly the one for an M/G/m/m system, which is a proof of property 3.31.

We also have (the proof is omitted)

Theorem: (5.10)
The response time for an M/G/1 simmetric queue is the same given in (5.11)

$$v(x) = \frac{x}{1-\rho} \tag{5.11}$$

Theorem: (5.12)
User i in a symmetric queue has already received an amount of serviceZ_i with pdf

$$f_Z(x) = \frac{1 - F_\chi(x)}{m_\chi}. \tag{5.13}$$

Furthermore, all those times are independent. The same holds for the amount of service user i has to receive yet.

We remark that the pdf in (5.13) is the one we already observed in the Renewal Paradox 2.28.

About the delay we can repeat for symmetric queues the results exposed at the end of Section **??**.

5.3 Priorities

In this section we show some results valid for an M/G/1 system with priorities. Users are grouped into classes with index p, $p = 1, 2, \ldots, P$ denoting its priority. Each class has its own average rate of arrivals λ_p and its own service time pdf $b_p(x)$. Arrivals enters P different queues, each one served with the FCFS policy. When a customer must be selected to enter service, it is taken from the highest priority non-empty queue.

The global characteristics of the arrivals are described by the following parameters

$$\lambda = \sum_{p=1}^{P} \lambda_p,$$
$$b(x) = \sum_{p=1}^{P} \frac{\lambda_p}{\lambda} b_p(x) \tag{5.14}$$

$$\rho_p = \lambda_p m_{xp}$$
$$\rho = \lambda m_x = \sum_{p=1}^{P} \rho_p$$

We want to derive the averages of waiting times $E[W_p]$ and time in the system $E[V_p] = E[W_p] + m_{xp}$

5.3.1 Non-preemptive systems

If all queues have asymptotic distributions, let denote

$$\sigma_p = \sum_{i=p}^{P} \rho_i, \qquad \sigma_1 = \rho.$$

We have **Theorem:** (5.15)
The average waiting time in the queue of priority p is given by:

$$E[W_p] = \frac{\rho E[Z]}{(1-\sigma_p)(1-\sigma_{p+1})} \qquad p = 1, 2, \ldots, P-1$$
$$E[W_P] = \frac{\rho E[Z]}{1-\rho_P} \tag{5.16}$$

where $E[Z]$ is, as usual, the average time to service completion. It can be derived by (5.14) or, as

$$E[Z] = \sum_{p=1}^{P} \frac{\rho_p}{\rho} E[Z_p] \tag{5.17}$$

where Z_p is the completion time of a class p customer.

Proof

We proceed in a way similar to the one used in proving the Pollaczek-Kinchin formula. The average waiting time in the queue of a client class p has three components,

$$E[W_p] = E[W_p^I] + E[W_p^{II}] + E[W_p^{III}]$$

The first is the completion time of the service found on arrival. Since this time is different from zero with probability ρ, we have:

$$E[W_\rho^I] = \rho E[Z]$$

** The second component is the waiting time caused by the service of users found in queues of priority p and greater. Denoted by N_i the number of users found in queue i, we have

$$E[W_p^{II}] = \sum_{i=p}^{P} E[N_i] m_{xi}$$

and, using Little's result $E[N_i] = \lambda_i E[W_i]$, becomes:

$$E[W_p^{II}] = \sum_{i=p}^{P} \rho_i E[W_i]$$

The third component is the waiting time caused by the service of users, of priority $p+1$ and greater, that arrive while the customer under exam is in the queue, i.e., during time W_p. Denoted by M_{ip} the number of such users belonging to the class i, we have:

$$E[W_p^{III}] = \sum_{i=p+1}^{P} E[M_{ip}] m_{xi}$$

Since $E[M_{ip}|W_p = w] = \lambda w$, by the Total Probability Theorem we have $E[M_{ip}] = \lambda_i E[W_p]$ (this is not Little's result, since $i \neq p$), and we can write

$$E[W_p^{III}] = \sum_{i=p+1}^{P} \rho_i E[W_p]$$

Taking the summation,

$$E[W_p] = \rho E[Z] + \sum_{i=p}^{P} \rho_i E[W_i] + \sum_{i=p+1}^{P} \rho_i E[W_p] = \rho E[Z] + \sum_{i=p+1}^{P} \rho_i E[W_i] + E[W_p]\sigma_p$$

and solving with respect to $E[W_p]$ yields

$$E[W_p] = \frac{\rho E[Z] + \sum_{i=p+1}^{P} \rho_i E[W_i]}{1 - \sigma_p}.$$

This is a system of equations that can be solved recursively starting from $p = P$, showing the thesis. ♣

We have so far assumed that all queues are in stationary equilibrium. However, it is possible that equilibrium does not hold for lower priority queues. This happens if there exists an r such that $\sigma_r > 1$ while $\sigma_{r+1} < 1$. In this case the first $P - r$ queues are in equilibrium, the others will have countless users. Queues with priority $1, 2, \ldots, r-1$ never get access to the service (asymptotically) and can be removed. The formulas above are still valid provided that we use

$$\rho_r = 1 - \sigma_{r+1} \tag{5.18}$$

$$\rho = \sigma_r = 1 \tag{5.19}$$

Example *Synchronous M/D/1 system* (5.20)
Here we mean a system with Poisson arrival and detreministic service of duration of D, but the beginning of the service itself is synchronized to time instants $t_k = kD$ (it is a typical case of a queue of a TDMA transmitter when messages have to wait for the assigned slot). It may happen that the customer enters the an empty system, but it must still await the time to enter service.

Some simple results can be obtained by observing that synchronization can be provided by a dummy traffic with lower priority, still with service time D, which completely saturate the lower priority queue. We use (5.16) with $P = 2$, and we get

$$E[W_2] = \frac{D}{2(1-\rho_2)}$$

Note that, with no synchronization ($\rho_1 = 0$), we have

$$E[W_2] = \frac{\rho_2 D}{2(1-\rho_2)}.$$

The time difference between the two cases is $D/2$, as could be expected.♣

5.3.2 Priority assignment

The average delay experienced by a user, regardless of its priority, is

$$E[W] = \sum_{i=1}^{P} \frac{\lambda_i}{\lambda} E[W_i]. \tag{5.21}$$

The question arises whether, maintaining the different user classes, the average delay $E[W]$ changes assigning priorities in a different order, and, if this is the case, which assignment provides the smallest delay $E[W]$.

To this purpose we prove the following

Theorem: (5.22)
The average delay of an FCFS system with different classes of users can be expressed as :

$$E[W]_{FCFS} = \sum_{i=1}^{P} \frac{\rho_i}{\rho} E[W_i]. \tag{5.23}$$

Proof
As we did to prove the average delays of the different classes, the average delay of an FCFS system can be written as

$$E[W]_{FCFS} = \sum_{i=1}^{P} E[N_i]m_{xi} + \rho E[Z] = \sum_{i=1}^{P} \rho_i E[W_i] + \rho E[Z].$$

The same delay can also be expressed by the pollaczek-Khinchine formula

$$E[W]_{FCFS} = \frac{\rho E[Z]}{1-\rho},$$

Substituting $\rho E[Z]$ from the above into the preceding one, and solving for $E[W]_{FCFS}$ we get the thesis. ♣

Now, $E[W]_{FCFS}$ (5.23) can not be altered by re-assigning the priorities of the different classes. Comparing with (5.21) we see that neither $E[W]$ can be altered if

$$\frac{\rho_i}{\rho} = \frac{\lambda_i}{\lambda},$$

which happens when $m_{xi} = m$, i.e., when the average service time of all classes are equal. Otherwise $E[W]$ can be different from $E[W]_{FCFS}$, and can change with the priority re-assignment. However, when priorities are assigned regardless of service times, all the priority classes have the same average service time and, again, we have $E[W] = E[W]_{FCFS}$. In order to get a change in $E[W]$ the re-assignment of priority classes must take into account the average service time, i.e., the selection policy must be Service Dependent. We have

Theorem: (5.24)
the priority assignment that minimizes $E[W]$ is the one that assigns priorities in reverse order with respect to the average service times.

Proof
Delay (5.21) can be rewritten as

$$E[W] = \frac{1}{\lambda} \sum_{i=1}^{P} \frac{1}{m_{xi}} \rho_i E[W_i] \tag{5.25}$$

The problem is to reassign the indexes i of of classes to indexes of priority p in such a way that the product of two functions, $1/m_{xi}$ e $\rho_i E[W_i]$, is minimal, knowing that $\sum \rho_i E[W_i]$, is constant (for the work conserving law (5.23)).

Let consider the class a with the lowest average service time, denoted by m_a, and the other classes together b, with average service time m_b. There are two possibilities: either the highest priority is assigned to a, in which case we have

$$\lambda E[W'] = \frac{1}{m_a}\rho_a E[W'_P] + \frac{1}{m_b}\rho_b E[W'_p] \tag{5.26}$$

or it is assigned to a class within group b, in which case we have

$$\lambda E[W"] = \frac{1}{m_a}\rho_a E[W"_p] + \frac{1}{m_b}\rho_b E[W"_P],$$

and the difference is

$$\lambda(E[W'] - E[W"]) = \frac{\rho_a}{m_a}\left(E[W'_P] - E[W"_p]\right) + \frac{\rho_b}{m_b}\left(E[W'_p] - E[W"_P]\right) \tag{5.27}$$

From the work-conserving law we write

$$\rho_a E[W'_P] + \rho_b E[W'_p] = \rho_a E[W"_p] + \rho_b E[W"_P],$$
$$\rho_a E[W'_P] - \rho_a E[W"_p] = \rho_b E[W"_P] - \rho_b E[W'_p] \leq 0 \tag{5.28}$$

where the inequality comes from the fact that $\rho_a E[W'_P] \leq \rho_a E[W"_p]$. From (5.28) we also have

$$\frac{\rho_a}{m_a}\left(E[W'_P] - E[W"_p]\right) = \frac{\rho_b}{m_a}\left(E[W"_P] - E[W'_p]\right) < \frac{\rho_b}{m_b}\left(E[W"_P] - E[W'_p]\right) \leq 0,$$

because $1/m_a > 1/m_b$, but the difference is negative. Finally, subtracting the third term to the first, we have

$$\frac{\rho_a}{m_a}\left(E[W'_P] - E[W"_p]\right) + \frac{\rho_b}{m_b}\left(E[W'_p] - E[W"_P]\right) < 0,$$

which proves (5.27) is negative, and the first assignment (5.26) minimizes $E[W]$.

Having assigned the highest priority, we repeat the same argument for the class with second least average service time, and so on, proving the thesis. ♣

5.3.3 Shortest-Job-First discipline

The results of the previous section show that the Shortest-Job-First (SJF) policy is the non-preemptive one that minimizes $E[W]$. Here we derive response times of this policy assuming that service time X is a continuous RV. We use (5.16) and divide the service x, $0 \leq x \leq T$ into $P = T/\Delta x$ classes, such that

$$(i-1)\Delta x \leq x < i\Delta x, \qquad 1 \leq i \leq P.$$

Then, priorities are assigned so that $p = P - i$, and we have

$$(p-1)\Delta x < T - x \leq p\Delta x, \qquad 1 \leq p \leq P.$$

In so doing, the highest priority $i = P$ corresponds to service $x \approx 0$ and the lowest, $p = 1$, to $x \approx T$. Turning to the continuous case, where $\Delta x \to 0$ and $i \to \infty$, we have $i\Delta x \to x$, and the priority class becomes $p = T - x$. Then we let $T \to \infty$. We have the following correspondences with (5.16):

$$\lambda_{T-x} \to \lambda b(x)dx; \qquad m_{T-x} \to x; \qquad \rho_{T-x} \to \lambda x b(x)dx;$$

$$\sigma_{p+1} \simeq \sigma_p = \sum_{i=p}^{P} \rho_i \to \int_0^x \lambda y b(y)dy.$$

Then, (5.16,) becomes

$$E[W/x] = \frac{\rho E[Z]}{\left(1 - \lambda \int_0^x y b(y)dy\right)^2}. \tag{5.29}$$

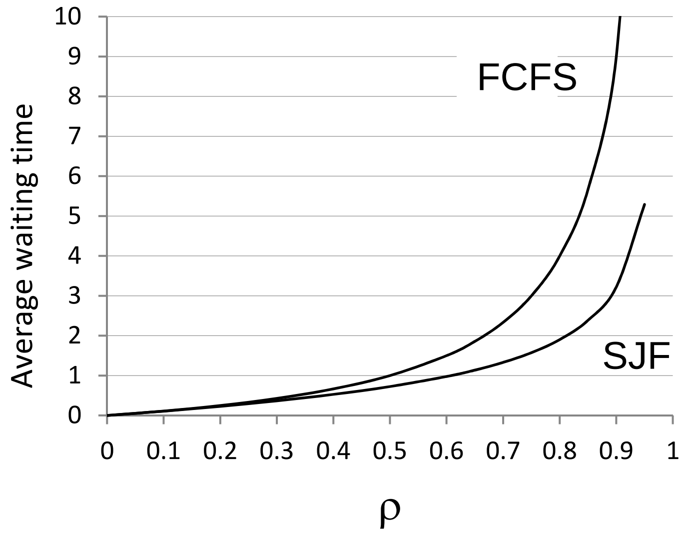


Figure 5.1: *Average waiting time versus ρ for two different disciplines, and negative exponential service time with mean 1.*

The above grows as x grows, and we have

$$\lim_{x\to 0} E[W/x] = \rho E[Z],$$

$$\lim_{x\to\infty} E[W/x] = \frac{\rho E[Z]}{(1-\rho)^2},$$

$$E[W] = \int E[W/x]b(x)dx.$$

The comparison of above expression with the FCFS case,

$$E[W]_{FCFS} = \frac{\rho E[Z]}{1-\rho},$$

is shown in Figure 5.1, which also compares the average waiting time with FCFS and SJF for M/M/1 mean service time 1.

Example (5.30)

Let consider a system with Poisson arrivals and uniform service time between 0 and 2T. Compare the gain, with respect to the average waiting time of the FCFS policy, of the following policies

a) *two priorities, the higher to users with service time between 0 and T;*

b) *Processor Sharing policy;*

c) *SJF discipline.*

Also compare the conditional waiting time $E[W/x]$.

For the uniform distribution in 0 e $2T$ we have

$$m_X = T \qquad \sigma_x^2 = \frac{T^2}{3} \qquad E[Z] = \frac{2}{3}T$$

$$E[W]_{FCFS} = \frac{2T\rho}{3(1-\rho)}$$

a) For services with priority 1 and 2 (service time between T and $2T$, and 0 and T respectively), we have :

$$E[W_1] = \frac{\rho E[Z]}{(1-\rho)(1-\rho_2)}$$

$$E[W_2] = \frac{\rho E[Z]}{1-\rho_2}$$

and, therefore

$$E[W] = \frac{\lambda_1}{\lambda_1+\lambda_2}E[W_1] + \frac{\lambda_2}{\lambda_1+\lambda_2}E[W_2]$$

Since it is

$$\frac{\lambda_1}{\lambda_1+\lambda_2} = \frac{\lambda_2}{\lambda_1+\lambda_2} = \frac{1}{2}$$

We finally get

$$E[W] = \frac{\rho(2-\rho)}{2(1-\rho_2)(1-\rho)}E[Z]$$

with

$$\rho = \lambda T, \qquad \rho_1 = \lambda_1 \frac{3}{2}T = \frac{3}{4}\rho, \qquad \rho_2 = \lambda_2 \frac{T}{2} = \frac{1}{4}\rho, \qquad E[Z] = (2/3)T$$

Reassuming, we have

$$E[W]_{2classi} = \frac{T\rho(2-\rho)}{3(1-\rho_2)(1-\rho)} = \frac{T\rho(2-\rho)}{3(1-\rho/4)(1-\rho)}$$

and the gain is

$$\eta_1 = \frac{E[W]_{2classi}}{E[W]_{FCFS}} = \frac{(2-\rho)}{2(1-\rho/4)} = \frac{2-\rho}{2-\rho/2} < 1$$

We also have

$$E[W/x] = \begin{cases} E[W_2] & = & \dfrac{2\rho T}{3(1-\rho/4)} & \text{se } 0 \le x \le T \\ \\ E[W_1] & = & \dfrac{2\rho T}{3(1-\rho)(1-\rho/4)} & \text{se } T < x \le 2T \end{cases}$$

b)With Processor Sharing policy we have

$$E[W]_{PS} = \frac{T\rho}{1-\rho}$$

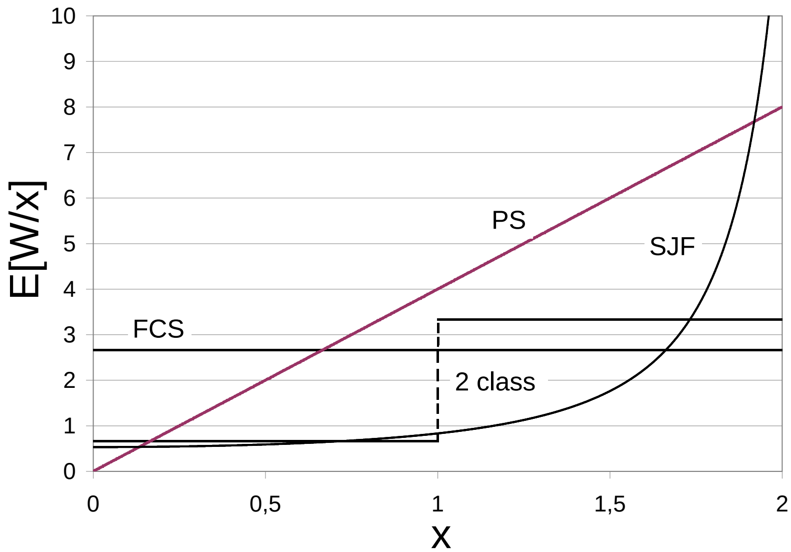


Figure 5.2: *Conditional average waiting time $E[W/x]$ as function of the requested amount of service x, with three different disciplines, for a uniform service time and $\rho = 0.8$.*

The gain is

$$\eta_2 = \frac{E[W]_{PS}}{E[W]_{FCFS}} = \frac{3}{2}$$

i.e., the PS policy present a higher average waiting time than FCFS. Om the oter side the Conditional waiting time is

$$E[W/x] = \frac{\rho}{1-\rho}x$$

c) Finally, with Shortest Job First policy we have

$$E[W] = \int E[W/x]b(x)dx = \int_0^{2T} E[W/x]\frac{1}{2T}dx$$

with

$$E[W/x] = \frac{\rho E[Z]}{\left(1 - \lambda \int_0^x yb(y)dy\right)^2} = \frac{2T\rho}{3\left(1 - \lambda\frac{x^2}{4T}\right)^2}.$$

After some algebra we get

$$E[W]_{SJF} = \frac{T\rho}{3}\left(\frac{\tanh^{-1}(\sqrt{\rho})}{\sqrt{\rho}} + \frac{1}{1-\rho}\right)$$

and the gain is

$$\eta_3 = \frac{E[W]_{SJF}}{E[W]_{FCFS}} = \frac{1}{2} + \frac{(1-\rho)\tanh^{-1}(\sqrt{\rho})}{2\sqrt{\rho}} \leq \eta_1 < 1.$$

Then, the SJF policy guarantee a smaller average waiting time with respect to both, FCFS and the two classes policy

Figure 5.2 shows the conditional average waiting time $E[W/x]$ as function of x, for FCFS and the three cases analyzed (with $\lambda = 0.8$ and $T = 1$).

5.3.4 Priority in Preemptive Systems

In these systems, an arrival occurring when a lower priority user is in service interrupts the latter and takes its place in the server. The preempted user gets back at the head of its queue and when in service again, it will resume service for the remaining part.

Theorem: (5.31)
The average waiting time in the queue of priority p is given by:

$$E[W_p] = \frac{m_{xp}(1-\sigma_p)\sigma_{p+1} + \sum_{i=p}^{P} \rho_i E[Z_i]}{(1-\sigma_p)(1-\sigma_{p+1})} \qquad p = 1, 2, \ldots, P-1 \tag{5.32}$$

$$E[W_P] = \frac{\rho_P E[Z_P]}{1-\rho_P}$$

Proof
Since a customer can be overtaken by those of higher priority that arrive during its time in the system V (and not just in W as in the non-preemptive case), it is convenient to derive time $E[V_p]$.

Again, we have three delay components for a user of priority p. The first is the user's service

$$E[V_p^I] = m_{xp}$$

The second delay component is the waiting time caused by the service of users found in queues of priority p and greater, since lower priority users are preempted, and do not cause delay. This time coincides with the waiting time $E[W]$ of a FCFS system with no priority classes, where we have dropped classes lower than p:

$$E[V_p^{II}] = \frac{\rho'}{1-\rho'} E[Z'] = \frac{\sigma_p}{1-\sigma_p} \sum_{i=p}^{P} \frac{\rho_i}{\sigma_p} E[Z_i]$$

The third component is similar to that in the non-preemptive case, i.e., the waiting time caused by the service of users, of priority $p+1$ and greater, that arrive while the customer under exam is in the system. Since the time in the system of such user is on the average $E[V_p]$, the average number of such users is

$$\sum_{i=p+1}^{P} \lambda_i E[V_p]$$

and the corresponding delay is

$$E[V_p^{III}] = \sum_{i=p+1}^{P} \rho_i E[V_p].$$

Therefore we have

$$E[V_p] = m_{xp} + \frac{\sum_{i=p}^{P} \rho_i E[Z_i]}{1-\sigma_p} + \sum_{i=p+1}^{P} \rho_i E[V_p],$$

and, solved for $E[V_p]$, provides

$$E[V_p] = \frac{m_{xp}(1-\sigma_p) + \sum_{i=p}^{P} \rho_i E[Z_i]}{(1-\sigma_p)(1-\sigma_{p+1})} \qquad p = 1, 2, \ldots, P-1$$
$$E[V_P] = m_{xP} + \frac{\rho_P E[Z_P]}{1-\rho_P} \tag{5.33}$$

Taking off the average service time we have the result. ♣

We can actually see that the highest priority class behaves as in a system without priority.

We have then the following property, whose demonstration is omitted,

Theorem: (5.34)
The number in the system $N(t)$, under the Shortest-Remaining-Job-First policy is such that

$$N(t) \leq N'(t), \qquad \forall t$$

where $N'(t)$ represents the same process under any other work-conserving policy.

From this theorem we see that the SRJF discipline is the one that minimizes $E[N], E[V], E[W]$.

5.4 Advanced Material

5.4.1 Response time constraints

Esistono altre discipline che cambiano la legge $v(x)$. In ogni caso, dal vincolo della costanza del lavoro, $E[L] = E[W]_{FCFS}$, derivano dei vincoli di conservazione sui tempi di risposta. Prima di vedere questi vincoli definiamo la funzione $s(\tau)$ come il servizio restante a distanza τ dall'ingresso del cliente che chiede servizio x (Figura 5.3a). Denominiamo poi *carico* del cliente l'area del servizio restante

$$S = \int s(\tau) d\tau \tag{5.35}$$

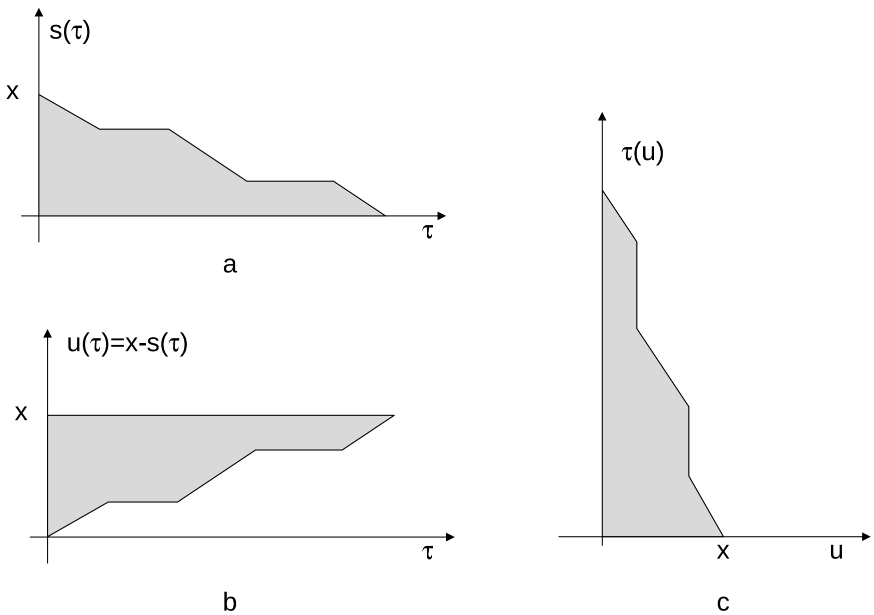


Figure 5.3: .

Dimostriamo ora un risultato simile al risultato di Little:

Lemma (5.36)
In un sisetma G/G il lavoro medio $E[L]$ presente nel sistema si può esprimere come

$$E[L] = \lambda E[S] \tag{5.37}$$

Proof
Dimostrazione molto simile a quella del risultato di Little. Consideriamo un ciclo di rigenerazione che comprende un Busy Period e consideriamo la funzione $l(t)$ lavoro presente nel sistema al tempo t. Per un processo rigenerativo ricorrente positivo la (1.204) assicura che si ha

$$E[L(t)] = \frac{E\left[\int_C L(t)dt\right]}{E[C]} \tag{5.38}$$

Detto S_i, $i = 1, 2, \ldots A$ il carico relativo al cliente i-esimo nel busy period si ha, deterministicamente

$$\int_C L(t)dt = \sum_{k=1}^{A} S_k \tag{5.39}$$

Sostituendo la (5.39) nella (5.38) si ha

$$E[L(t)] = \frac{E\left[\int_C L(t)dt\right]}{E[C]} = \frac{E\left[\sum_{k=1}^{A} S_k\right]}{E[C]} = \frac{E[A]}{E[C]}E[S] = \lambda E[S].$$

dove l'ultima eguaglianza viene dall'applicare l'uguaglianza (5.38) al processo tempo discreto degli arrivi (l'integrale diventa la somma di valore A).♣

Siamo ora in grado di mostrare il seguente teorema di conservazione:

Theorem: (5.40)
In un sistema G/G il lavoro medio nel sistema può essere espresso come

$$\lambda \int_0^\infty v(x)(1 - B(x))dx = E[L] \tag{5.41}$$

Proof
Come si vede dalle figure 5.3, il carico può anche essere calcolato come

$$S = \int_0^x \tau(u)du$$

che fornisce come media condizionata

$$E[S/x] = \int_0^x E[\tau(u)]du = \int_0^x v(u)du$$

dove si è riconosciuto che la funzione integranda è il tempo di risposta. Mediando sul tempo di servizio richiesto si ha

$$E[S] = \int_0^\infty \left(\int_0^x v(u)du\right) b(x)dx = \int_0^\infty \left(\int_0^\infty v(u)o(u,x)du\right) b(x)dx = \int_0^\infty v(u)(1-B(u))du$$

dove $o(u, x)$ è una funzione che vale 1 per $u \leq x$. Applicando il Lemma 5.36 si dimostra la 5.41.♣

Il vincolo mostra che se si riduce $v(x)$ per x piccoli, esso deve alzarsi per x grandi.

Si può anche mostrare che, per sistemi con arrivi di Poisson il lavoro medio in coda può anche essere scritto come

$$m_X \int_0^\infty w(x)(1 - B(x))dx = E[L] \tag{5.42}$$

Nella disciplina *Time Sharing* i clienti ricevono solo un quanto di servizio Δx alla volta e poi tornano in coda, se non lo hanno completato. Anche qui la disciplina della coda può essere varia, se si eccettua il caso in cui il cliente torna in testa alla coda, che coincide con il servizio FCFS senza preemption.

Attenzione particolare merita il caso Time sharing quando Δx è infinitesimo. Questo caso il comportamento del sistema è indistinguibile dal caso Processor Sharing dal momento che, se il sistema è stabile (numero in coda finito), anche il tempo in coda in attesa del prossimo quanto di servizio è infinitesimo e dunque in un tempo finito, tutti i clienti ricevono la stessa porzione di servizio. Tutto funziona esattamente come se tutti i clienti in coda, in numero di i, stessero ricevendo il servizio a velocità i volte più piccola. Anche in questo caso valgono tutte le considerazioni svolte per il caso Processor Sharing.

Le discipline considerate finora sono tali da avvantaggiare nello stesso modo i clienti con tempo di servizio breve. Considerando la processor-sharing, un nuovo cliente riceve immediatamente servizio, ma solo una frazione uguale agli altri. Se si modifica il servizio in modo che il cliente arrivato riceva *totalmente* il servizio in modo dedicato finchè il servizio ricevuto non eguaglia quello ricevuta da alcun altro cliente, clienti con servizi corti sono ancor più avvantaggiati.

La disciplina modificata come sopra prende il nome di *Foreground-Background*. Inoltre, quando esistono clienti con servizio ricevuto uguale, e non ne esistano con servizio ricevuto minore, la procedura applicata a questi è la processor sharing.

Il calcolo di $v(x)$ per questa disciplina può essere fatto in modo semplice osservando che su questo non influenzano i servizi che durano più di x. Infatti il sistema può essere concepito come due stadi in cascata dove nel primo ci sono coloro che hanno ricevuto un servizio minore o uguale a x e nel secondo gli altri. Il cliente che deve ricevere x non entrerà nel secondo ma, cosa più importante, il servente non abbandona il primo stadio, che serve esclusivamente, e lo fa solo quando anche tutti i clienti presenti lo lasciano, avendo ricevuto x. Il tempo $v(x)$ è il tempo di attraversamento del primo stadio che si calcola facilmente data che devono uscire tutti contemporaneamente. Infatti, in questo caso, tale tempo è pari al totale lavoro trovato all'arrivo più il lavoro portato più il lavoro arrivato durante la permanenza (che dura x):

$$v(x) = L'_x + x + L''_x$$

Il lavoro L'_x è quello trovato in un sistema FCFS in cui il tempo di servizio abbia d.d.p.

$$b_x(z) = b(z) + \delta(z-1) \int_x^{\infty} b(z)dz \qquad x \leq z \tag{5.43}$$

Detto $m(x)$ il valor medio della (5.43), il lavoro L''_x è pari a $m(x)$ per il numero di arrivi durante x e dunque si ha

$$v(x) = L'_x + x + \lambda v(x) m(x)$$

che, posto $\rho_x = \lambda m(x)$, risolta fornisce

$$v(x) = \frac{L'_x + x}{1 - \rho_x} \tag{5.44}$$

Il valore della (5.44) nell'origine vale 0 mentre la sua pendenza vale 1, contrariamente al caso Processor Sharing dove vale $1/(1-\rho)$. Dunque, i servizi molto corti sono avvantaggiati rispetto al Processor Sharing, a scapito di quelli più lunghi. Per il tempo medio complessivo si ha

$$E[V] = \int \frac{L'_x}{1-\rho_x} b(x)dx + E[V]_{PS} \tag{5.45}$$

avendo indicato con $E[V]_{PS}$ il tempo medio d'attesa nei sistemi Processor Sharing e analoghi. $E[V]$ è dunque in generale maggiore che nei sistemi citati e ciò indica che il sistema potrebbe comportarsi male nel caso i clienti non abbiano tempi di servizio molto diversi fra loro.

5.4.2 Sistema GI/G/1 con gestione LCFS preemption-resume

Purtroppo in un sistema GI/G/1, il processo $N(t)$ non è studiabile analiticamente, e nemmeno con il metodo usato per il sistema M/G/1. Infatti il processo $N(t_j^+)$ non è markoviano.

Alcuni risultati possono essere invece ricavati per un sistema GI/G/1 con interruzione. E' un sistema in cui il cliente che arriva entra immediatamente nel servizio, facendo cessare colui che eventualmente vi si trova. Quest'ultimo riprende il servizio nel punto in cui è stato interrotto, solo dopo che tutti quelli arrivati dopo di lui sono stati serviti e sono usciti dal sistema.

Questo sistema è relativamente semplice perchè gli istanti di arrivo dei clienti sono istanti di rigenerazione relativi. Infatti tutto quello che accade nel tempo fra l'arrivo e l'uscita di uno stesso cliente, non dipende da ciò che è successo prima del suo arrivo, né da ciò che succederà dopo la sua partenza.

Chiamiamo tale intervallo di tempo $B_i^{(k)}$ (Busy period di livello i) dove k è un numero d'ordine (sono tutti equidistribuiti) e i è il minimo numero di clienti nel sistema durante tale periodo (si veda la Figura 5.4). B_1 è così il Busy period dell'intero sistema. Poichè quando arriva un cliente tutto si rinnova, il comportamento del processo $(N(t)-i+1)$ durante questi cicli B_i *non dipende da* i.

Si noti che si può scrivere:

$$B_i = X_i + \sum_{k=1}^{A_i} B_{i+1}^{(k)} \tag{5.46}$$

dove X_i è la durata del servizio del cliente che inizia il ciclo B_i, e A_i è il numero di arrivi, cioè il numero di volte che tale servizio è interrotto, cn la convenzione che la sommatoria vale zero se $A_i = 0$. Tutte le grandezze sopra non dipendono da i. Inoltre le V.C. B e A sono incorrelate. Posto dunque $E[A] = \sigma$, si ha:

$$E[B] = m_x + \sigma E[B] \tag{5.47}$$

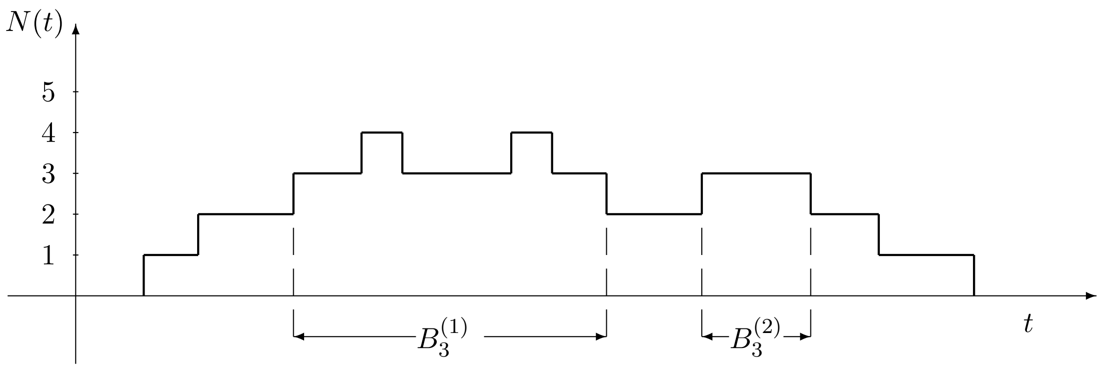


Figure 5.4:

da cui si ottiene

$$E[B] = \frac{m_x}{1-\sigma}$$

Ricorrendo al solito Teorema **??** per le distribuzioni asintotiche delle catene rigenerative, si ha:

$$P(N \geq i+1/N \geq i) = \frac{P(N \geq i+1)}{P(N \geq i)]} = \frac{\frac{E[A_i]E[B_{i+1}]}{E[C]}}{\frac{E[B_i]}{E[C]}} = E[A_i] = \sigma \qquad i = 1, 2, \dots$$

Dalla relazione sopra e dal fatto che si sa che $P(N \geq 1) = \rho = \lambda m_x$ si ottiene facilmente

$$\begin{aligned} \pi_i &= \rho(1-\sigma)\sigma^{i-1} \qquad\qquad i = 1, 2, \dots \\ \pi_0 &= 1 - \rho \end{aligned} \tag{5.48}$$

$$E[N] = \frac{\rho}{1-\sigma}$$

L'equazione (5.48) ci indica una soluzione abbastanza semplice, ma purtroppo non è possibile ricavare in generale il valore di σ.

Cerchiamo ora la distribuzione all'ingresso q_i. Partiamo dal fatto che:

$$\frac{\pi_{i+1}}{\pi_i} = \sigma \qquad\qquad i = 1, 2, \dots$$

e, sfruttando la relazione di equilibrio dei sistemi di nascita e morte generalizzati, otteniamo

$$\frac{\lambda_i^*}{\mu_{i+1}^*} = \sigma \qquad\qquad i = 1, 2, \dots \tag{5.49}$$

Precedentemente s'è detto come le caratteristiche statistiche del processo $N(t)$ non dipendano dal livello $i > 0$ cui si fa riferimento. Ciò comporta che λ_i^* e μ_i^* non dipendano, per $i > 0$, da i stesso. Ma allora si hanno le :

$$\mu_i^* = 1/m_x$$
$$\lambda_i^* = \sigma/m_x \qquad\qquad i = 1, 2, \ldots$$

Siamo in grado di trovare l'ultimo termine ignoto λ_0^* sfruttando il fatto che $\lambda = \sum \pi_i \lambda_i^*$ e che $\pi_0 = 1 - \rho$, e risulta essere:

$$\lambda_0^* = \lambda \frac{1-\sigma}{1-\rho}$$

Il coefficente $\sigma = \lambda^* m_x$ viene chiamato fattore di utilizzo specifico. I risultati ottenuti, insieme al Teorema **??**, permettono infine di scrivere che:

$$q_i = (1-\sigma)\sigma^i \tag{5.50}$$

Nel caso di arrivi di Poisson (M/G/1), si riconosce facilmente che

$$\sigma = \lambda m_x = \rho$$

Dunque la distribuzione vista da chi arriva è geometrica, come nel caso M/M/1 senza interruzione. Con servizi generali, ciascuno dei clienti in coda ha ancora un lavoro pari a Z da fare e il lavoro totale nel sistema eguaglia il tempo d'attesa in un sitema FCFS senza interruzione. Risulta dunque spiegato il significato dell'espressione (**??**).

In un sistema M/M/1, la coincidenza delle distribuzioni, con e senza interruzione, si spiega osservando che, l'interruzione di un servizio esponenziale da parte di chi arriva, lascia un servizio da completare che è ancora esponenziale con lo stesso valor medio (proprietà di non memoria).

Quanto appena detto, vale però per qualunque sistema con servizio esponenziale (GI/M/1) per i quali, dunque, le distribuzioni (5.48) e (5.50) valgono anche per un sistema non interruttivo. In quest'ultimo caso anzi, il parametro σ è determinabile e si può mostrare che è l'unica soluzione compresa fra 0 e 1 dell'equazione:

$$\sigma = A^*((1-\sigma)/m_x)$$

dove $A^*(s)$ è la trasformata di Laplace della d.d.p. del tempo di intercorrenza fra gli arrivi.

Chapter 6

Network Models

6.1 Networks and Routing

A network can be represented by a collection of nodes connected by links. In networks there are entities (eg. vehicles, data packets, users) that at each node (junction, router, switch) must be routed toward an output (port) to proceed along their path from a source node to a destination node. We can distinguish among *open networks*, where users enter the network at some nodes and leave the network at some others, while the number of users inside the network changes as they keeps enetering and leaving the network. In *closed networks* users con neither enter the network nor leave, and their number M is constant as they keep recirculating inside the network itself.

Usually we assume to know the entering rate λ of new users in open networks, their number in closed netorks, and the behavior of each node with respect to routing and, in some cases, the processing time. In return, we want to know whether the operation of the nework can reach stationary conditions, and quantities such as flow rates λ_{ij} between nodes i and j,, the average number at each nodes and average delays, i.e. the average time it takes to cross network subsets (subnetwork).

In the following we assume that all users behave in the same probabilistic way, so that we need not distinguish among them. later on this assumption will be partially released.

6.1.1 The routing chain

Since all users are driven by the same probabilistic rules, we can define an integer process $X(n)$ (the routing process) that describes the node $x(n)$ reached, by a sample user, at routing step n (figure 6.2). The state space of $X(n)$ is represented by the set of the network nodes and, Vice-versa, the state space of any discrete-state process $X(n)$ can be seen as a network. The two concepts are interchangeable, and using one name or the other is only a matter of preference, possibly dictated by the application.

In networks, the one-step transition probabilities, p_{jk} of $X(n)$ are called routing probabilities, which

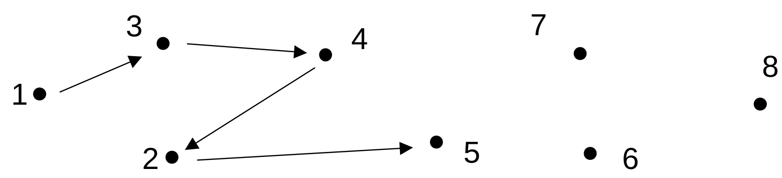


Figure 6.1: *Example of routing.*

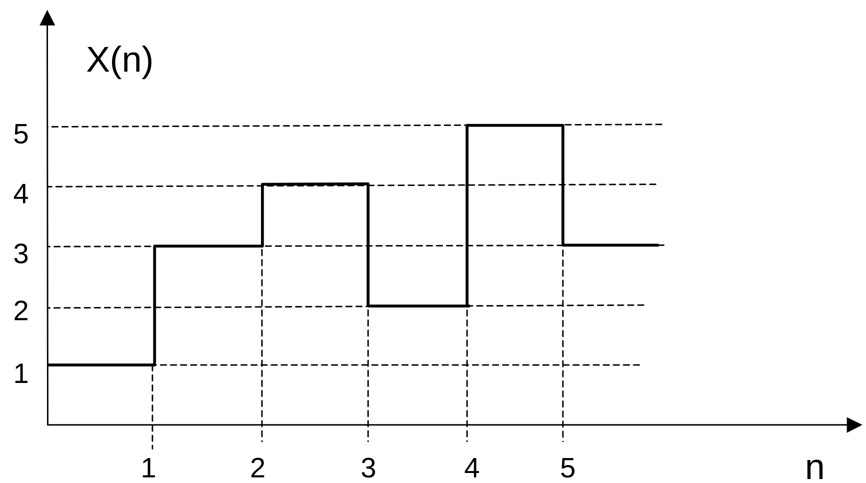


Figure 6.2: *Sample of the routing process $X(n)$ of figure 6.1.*

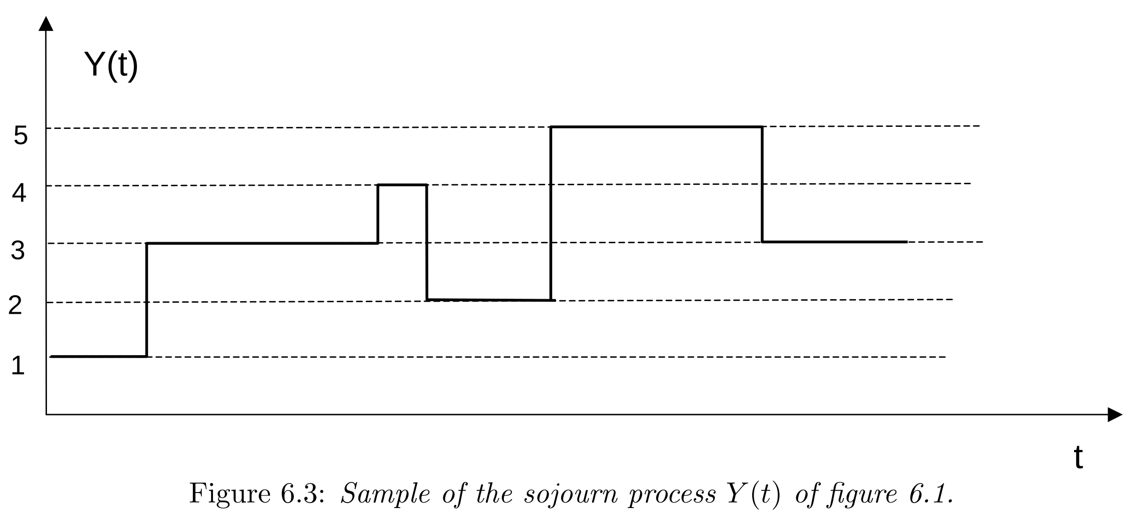


Figure 6.3: *Sample of the sojourn process $Y(t)$ of figure 6.1.*

we denote hereafter by ℓ_{jk}. The corresponding transition matrix is called the routing matrix.

We have seen in the preceding chapters that in stationary conditions, as we are always assume here, the routing probabilities are related to the probability distribution, here denoted by ν_j, by the balance of probabilistic fluxes (1.148), which is written, with the new symbols, as

$$\nu_i = \sum_{k=0}^{J} \nu_k \ell_{ki}, \tag{6.1}$$

where J is the number of nodes within the network. Here ν_i represent the probability that a "sample" users, visiting nodes according to the above routing matrix, is found in node i at routing step n, asymptotically.

If $X(n)$ is regenerative when entering in node 0, as it happens for markov chains, then ν_i may be expressed as (Theorem 1.189):

$$\nu_i = \frac{e_i}{\sum_j e_j}, \tag{6.2}$$

being e_j the number of times (in fact the amount of time) the process spent in node i in a cycle. Since it enters one time only in node 0, the above can also be expressed as

$$\nu_i = \nu_0 e_i. \tag{6.3}$$

Among the different routing processes $X(t)$ we can have the following:

Example *Purely Random Routing* (6.4)
Here, at each node i the routing to other nodes k is decided in a purely random way, irrespective of the current position and the past history, with probability ℓ_k. This case is obviously a very special case of no relevance dealing with networks.

Example *Markovian Routing* (6.5)
Here, at each node i the routing to other nodes k is decided in a purely random way, depending on the present current position i, but irrespective of the past history, with probability ℓ_{ik}. (Note the analogy with the random walk, of which the routing process is a generalization).

Example *Deterministic Routing* (6.6)
Here, the path in the network is fixed. We can distinguish two cases. If the routing network has no loops (no node is visited more than once), then the routing is still Markovian, although of a very specific case. Otherwise, if a node can be visited a specific number of times, memory takes place and the process is no longer markovian. Note also that, in the former case, we can have more than one path, with different origins, within the network only if these paths do not intersects, otherwise additional memory is introduced.

6.1.2 The sojourn chain

If we take into account the sojourn time in each node we get the continuous-time process $Y(t)$ (figure 6.3), defined as the node visited at time t by the sample user, and called the *sojourn chain*. Here transition (routing) probabilities are replaced by transition (routing) rates q_{jk}, and the probability distribution γ_i is attained by the balance equations

$$\gamma_i = \sum_{k=0}^{J} \gamma_k q_{ki}. \tag{6.7}$$

Distribution ν_i and γ_i are related, since the routing chain $X(n)$ can be seen as the transition chian of $Y(t)$ (see Section 1.15), and we have

$$\ell_{jk} = \frac{q_{jk}}{\sum_{j \neq k} q_{jk}}. \tag{6.8}$$

On the other side the corresponding of (6.2) is (Theorem 1.189):

$$\gamma_r = \frac{E[Y_r]}{\sum_i E[Y_i]},$$

and since $E[Y_r] = e_r E[V_r]$, where we have denoted by V_r the sojourn time in r, we have also, by (6.3),

Property (6.9)
distributions $\{\gamma_r\}$ and $\{\nu_r\}$ are proportional, and related by

$$\gamma_r = \frac{\nu_r E[V_r]}{\sum_i \nu_i E[V_i]}. \tag{6.10}$$

6.1.3 Network flows

Let consider the flow of users routed from node i to node k and let λ_{ik} the asymptotic rate of this flow. Similarly, λ_i is the rate of the flow of users that crosses node i (in stationary conditions equal to the rate at which users enter and leave node i). Clearly we have

$$\lambda_i = \sum_k \lambda_{ik}, \tag{6.11}$$

and

$$\lambda_i = \sum \lambda_{ki}. \tag{6.12}$$

Since we have also

$$\lambda_{ki} = \lambda_k \ell_{ki}, \tag{6.13}$$

relation (6.12) can also be rewritten as

$$\lambda_i = \sum_k \lambda_k \ell_{ki}. \tag{6.14}$$

We note that relation (6.14) is the same as (6.1), meaning that

Theorem: *basic* (6.15)
User rates λ_i are proportional to the distribution ν_j of the routing chain, i.e.

$$\lambda_i = K\nu_i. \tag{6.16}$$

By (6.13) the same proportion occurs between λ_{ik} and probability fluxes ν_{ik} of the routing chain.

The result above is not surprising since the flow rate can be expressed as

$$\lambda_{ik} = E[M]\gamma_i q_{ik}, \tag{6.17}$$

where $E[M]$ is the average number of users in the network and $\gamma_i q_{ik}$ is the flow rate due to a sample user. Taking the summation we have

$$\lambda_i = \sum_{k\neq i} \lambda_{ik} = E[M]\gamma_i \sum_{k\neq i} q_{ik} = E[M]\gamma_i(-q_{ii}) = E[M]\frac{\gamma_i}{E[V_i]} = K\nu_i, \tag{6.18}$$

where the last passage comes from (6.10).

Therefore, being the (6.17) proportional to a probability flux, the equality of (6.11) and (6.12) only reflects the fact that it represents the balance of probability fluxes into and out state i.

6.1.4 Open Networks

A network can be open or closed. An open network can be derived from a closed network where node 0 represents the source and the sink of users' flow λ_0 into and out of the network. If the flow λ_0 injected into the network is known, then all flow rates in the net can be derived according to the following

Theorem: (6.19)
User rate λ_i entering node i of an open network is proportional to the flow λ_0 in the following way

$$\lambda_i = \frac{\lambda_0}{\nu_0}\nu_i, \tag{6.20}$$

and is obtained as the unique solution of the following system of equations

$$\lambda_i = \ell_{0i}\lambda_0 + \sum_{k=1}^{J} \ell_{ki}\lambda_k, \qquad i = 1, \ldots, J. \tag{6.21}$$

Proof
System (6.21) is equivalent to system (6.14) where λ_0 is made explicit. This is a balance system of equations (equal to (6.1)) whose solution can be expressed as proportional to λ_0 (that correspond to probability ν_0), which in open systems is known as it represents the rate entering the network. This also prove that λ_i and ν_i are proportional, as in (6.20), that we have already seen in theorem 6.15. The constant is derived by writing (6.20) for flow λ_0.♣

Example (6.22)
In the network in Figure 6.4 the birth rate is $\lambda_0 = 10\ s^{-1}$. The births are then allocated between

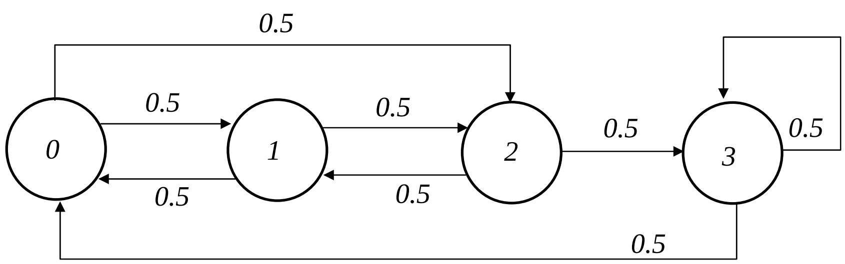


Figure 6.4: Example of a network.

node 1 and node 2 as shown in the Figure. The theorem tells us that the flows in the nodes are all equal because the solution of the routing chain is uniform. So $\lambda_i = 10\ s^{-1}$ for each i (note that flow λ_3 includes the loop, while that excluding the loop is halved. This also allows us to determine flows $\lambda_{ij} = \lambda_i \ell_{ij} = 5\ s^{-1}$. ♣

6.1.5 Little's Result Extended

The Little's result applies to nodes in the network and to any subnetwork as well, i.e., to any subset of nodes of the network. However, the following extension holds.

Theorem: (6.23)
The time Y_i the user spends in node $i \neq 0$ during its life (i.e., between its entering and exiting the open network), the number of customers N_i in node i and the rate λ_0 entering the network are related by

$$E[N_i] = \lambda_0 E[Y_i]. \tag{6.24}$$

Proof
Denoted by e_i the number an user enters node i in its life, we have

$$\lambda_i = e_i \lambda_0.$$

Then, the Little's Result applied to node i provides

$$E[N_i] = e_i \lambda_0 E[V_i].$$

The thesis is proven by observing that we also have

$$E[Y_i] = e_i E[V_i]. \quad \clubsuit \tag{6.25}$$

The result just shown allows us to provide an alternative interpretation of the distribution $\{\gamma_i\}$ of process $Y(t)$.

Corollary (6.26)
The asymptotic distribution of the sojourn process $Y(t)$ can be expressed as a function of the number of customers N_i in the generic node i as:

$$\gamma_i = \frac{E[N_i]}{\sum_{k=1}^{J} E[N_k]}. \tag{6.27}$$

Proof
The thesis comes immediately from the previous theorem applied to the fundamental Theorem 1.189.

6.1.6 Closed Networks

In closed systems, users do not enter nor leave the network, and their number M is constant. Therefore, the rates of the users' flows can not be obtained by the sole system (6.14) or (6.21) since λ_0 does not exists. However, we know that flows λ_i are proportional to probability ν_i (Property 6.15). We know also that they are proportional to the number of users in the system M. The proportionality constant is given by the following

Theorem: (6.28)
The flows λ_i are proportional to the distribution elements ν_i of the routing chain in following manner:

$$\lambda_i = \frac{\nu_i M}{\sum_{k=1}^{J} \nu_k E[V_k]}, \tag{6.29}$$

where M represents the number of users in the net.

Proof
Consider a section before node 0, or any other in the network. The flow rate λ_0 leaving the section can be seen as the flow into the network, but also as an outflow from the network (entering the section). For the Little's Result applied to the entire network we have

$$\lambda_i = \frac{M}{\sum_{k=1}^{J} E[Y_k]}.$$

The thesis comes from (6.25) through the relation (6.3)

$$e_k = \nu_k/\nu_0. \quad \clubsuit$$

Example (6.30)
Let consider a closed network with two nodes and M users. Users migrates back and forth to the

two nodes and the sojourn time are of two units in the first node and one unit in the second node, for each user independently from others.

We can derive $\lambda_1 = \lambda_2$ directly by observing that the time to complete a period equal 3 time units. Therefore $\lambda_1 = \lambda_2 = M/3$. This can also be derived by (6.29). ♣

Expression (6.29) shows that flow rates increase proportionally to M; they also decrease when sojourn times V_j increase, as it easily understood thinking of two nodes only. This does not happen with open networks, where flow rates only depends on λ_0 and not on sojourn times. This difference in the determination of flow rates can be explained observing that (6.29) holds also for open networks, where M must be replaced with its average value $E[M]$. Here, an increase in V_i is exactly counterbalanced by an increase in $E[M]$, due to the Little's result. Things become more complicated when dealing with closed network of queues, as shall see in later sections, where times V_i are functions of rates λ_i.

6.1.7 Network Delay

We call network delay the time it takes a user to cross the network, from its birth at node 0 to its death.

Theorem: (6.31)
The average network delay D is related to the average sojourn time at nodes by

$$D = \sum_i \frac{\lambda_i}{\lambda_0} E[V_i]. \tag{6.32}$$

Proof
The thesis comes immediately from Little's theorem applied to the network:

$$D = \frac{E[N]}{\lambda_0} = \frac{\sum_i E[N_i]}{\lambda_0},$$

and then using again Little's theorem to the nodes. Alternatively, we recognize that λ_i/λ in (6.32) represents the average number of times the user crosses node i. ♣

The results attained so far do not assume any specific process model for the flows entering or flowing in the network. They are valid even for a general routing chain, provided that we know the asymptotic values of ℓ_{ik}. The (6.32) is a fundamental formula in the field of telecommunication networks where constitutes a performance measurement. In general, the choice of the network route is carried out so as to minimize the network delay, that takes into account also the queues at the nodes.

Unfortunately, in nodes with queues, times $E[V_i]$, which in turn depend on λ_i, can be derived in the markovian case, that require a markovian routing mechanism and, in open networks, Poisson arrivals into the system. Even in the latter case, flows are altered by the service and queueing processes, they split and merge, and this may change the Poisson nature of flows. In the following we investigate under which circumstances flows out of the queues can still be considered Poisson.

6.2 Queue Interactions

When the nodes of a net are represented by queueing systems, the node sojourn time, and therefore the network delay, greatly depends on the input process. In a network of queues, the input of one is composed, at least in part, by the output of another. The Burke's theorem, together with union and the splitting properties of a Poisson flows assures that

Property (6.33)
In an open network with Poisson births, markovian routing, negative exponential servers with infinite waiting room, and with no closed paths, all the queues behave as if they were M/M systems in isolation.

In the above, markovian routing without closed paths and infinite waiting room is the only routing that assures a random splitting of the output flow, thus preserving the Poisson characteristic of each split flow. If we refer to closedpaths, even with the assumptions above, we still have a multi-dimensional markovian system, but we are forced to consider the nodes jointly, and the system state becomes the vector that describes the contents of nodes.

In the next subsections we discuss some simple examples, by means of state-space diagrams, that bring to light some properties about joint distribution and flows. In later sections those properties are generalized to broad classes of queuing networks.

6.2.1 Queues in parallel.

Let consider two M/M/1 systems with independent servers and independent arrivals with parameters $\lambda_1, \lambda_2, \mu_1, \mu_2$. We want to study the joint occupancy process $(N_1(t), N_2(t))$ by the two-dimensional Markov chain shown in Figure 6.5.

On the other hand, the independence of the arrival and service processes tells us that also the occupancy of the two queues are independent. This means that the joint distribution is the product of the two marginal distributions. We therefore have

$$\pi_{ij} = \pi_i^{(1)} \pi_j^{(2)} = (1-\rho_1)(1-\rho_2)\rho_1^i \rho_2^j \tag{6.34}$$

with $\rho_1 = \lambda_1/\mu_1$ e $\rho_2 = \lambda_2/\mu_2$.

It is left to the reader to verify that the above solution ensures the balance of flows. This, on the other hand, is implicit in the fact that along the columns only π_j changes and that this, as a solution of a Birth and Death process, ensures the balance of the vertical flows in both directions. Similarly for the rows and then the balance to nodes is guaranteed by the distribution above, that is therefore the solution.

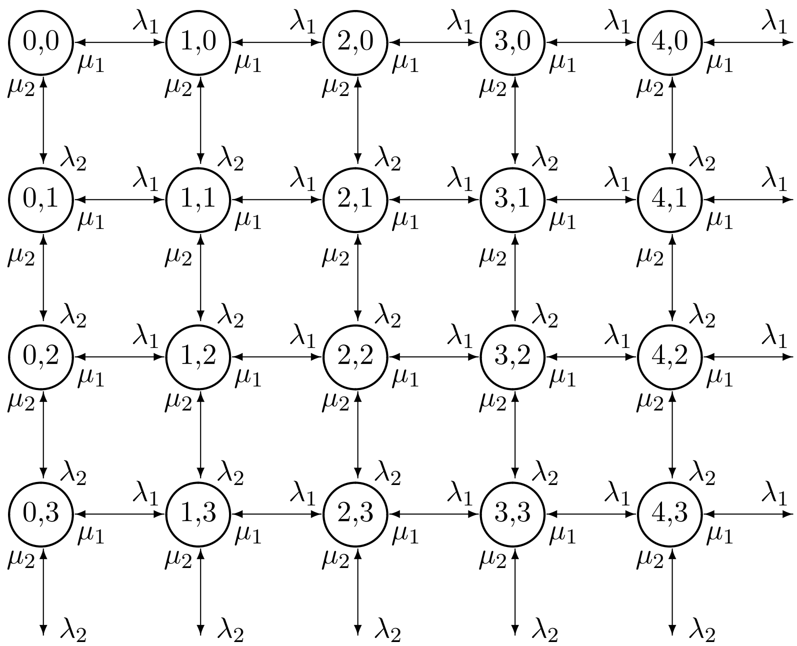


Figure 6.5: *State diagram of two queues in parallel.*

6.2.2 Open cascaded queues

Assume that the queueing systems of the previous case are cascaded, meaning that a Poisson flow of rate λ enters the first queue, while the departures from this queue enters the second one. We study the joint process $(N_1(t), N_2(t))$ through the two-dimensional Markov chain shown in Figure 6.6.

The solution of such a chain seems quite involved. However, it is possible to guess the solution from some geometric properties, and then we can check it into the balance equations. If these are satisfied, then we have found the solution, being it one and only one. Our guess is that the balance of probability fluxes is attained in the special pattern that sees the same fluxes along the closed meshes of type $(i, j) - (i + i, j) - (i, j + 1) - (i, j)$, i.e.,

$$\pi_{i+1,j}\mu_1 = \pi_{i,j+1}\mu_2 = \pi_{ij}\lambda, \tag{6.35}$$

If the above is true the balance at any node is attained by balancing the fluxes two by two, e.g., $\pi_{i,j+1}\mu_2 = \pi_{ij}\lambda$. Furthermore, the diagonal flux $\pi_{i+1,j}\mu_1$ can be replaced by flux $\pi_{i+1,j}\mu_1$ toward state (i, j) and by flux $\pi_{i,j}\lambda$ toward state $(i, j + 1)$. This replacement, carried out in each mesh, provides the diagram in Figure 6.7. In this figure the diagram is exactly the one of queues in parallel shown in Figure 6.7, which is balanced by the distribution

$$\pi_{ij} = (1 - \rho_1)(1 - \rho_2)\rho_1^i \rho_2^j, \qquad \rho_1 = \lambda/\mu_1, \quad \rho_2 = \lambda/\mu_2. \tag{6.36}$$

Alternatively, we can directly verify that (6.61) is satisfied by distribution (6.36). Hence, (6.36), strange as it may be, is indeed the asymptotic distribution of the cascaded queues.

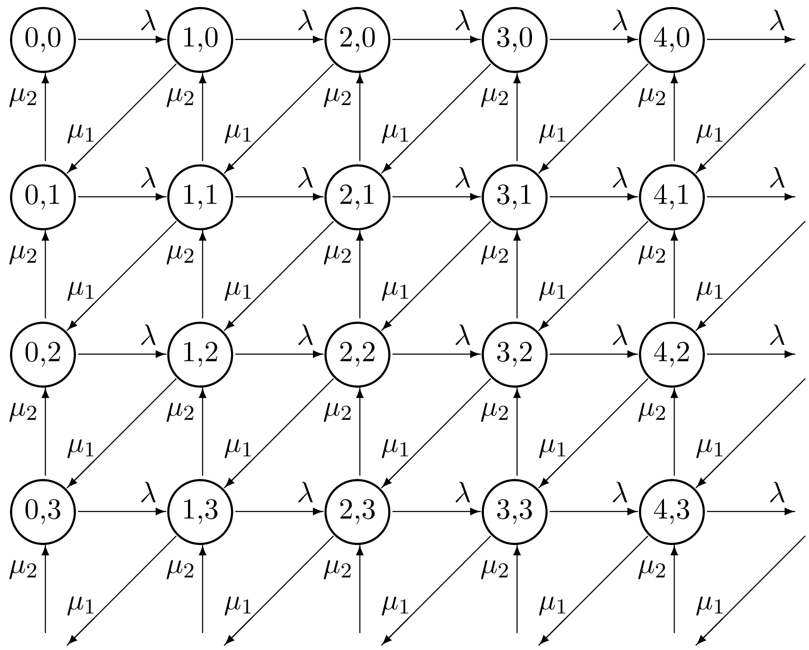


Figure 6.6: State diagram of two cascaded queues.

We may conclude, therefore, that at the same time t, the occupancies of the two systems are statistically independent. This may seem rather strange, since the occupancy of the two queues appear to strongly depend one of another. The fact is that this dependance appears only at different times, not at the same time.

The result attained above can be easily generalized to any number of servers, and the results still show the independence of the queues

6.2.3 Finite Capacity

Consider the cascaded queues where the second system has finite capacity finite Q, where the customers that end the service in the first system and find the second one full leave the network.

Referring to the diagram in Figure 6.6, this means that states $(i, Q+j), j > 0$ are dropped, together with their fluxes, whereas, fluxes of rate μ_1 out of states $(i, Q), i > 0$ is closed into states $(i-1, Q), i > 0$. This operation does not alter the balance of the fluxes of the infinite chain, which means that the solution is the same but the normalization constant.

If the finite queue belong to the first system and users that find this full leave the network, the corresponding dropping of states and fluxes alter the balance of the fluxes of the infinite chain, and the solution changes. This correspond to the fact that users flow that leave the first system is no longer Poisson.

On the other side, if the users that can not enter the first queue reach the second queue and leave

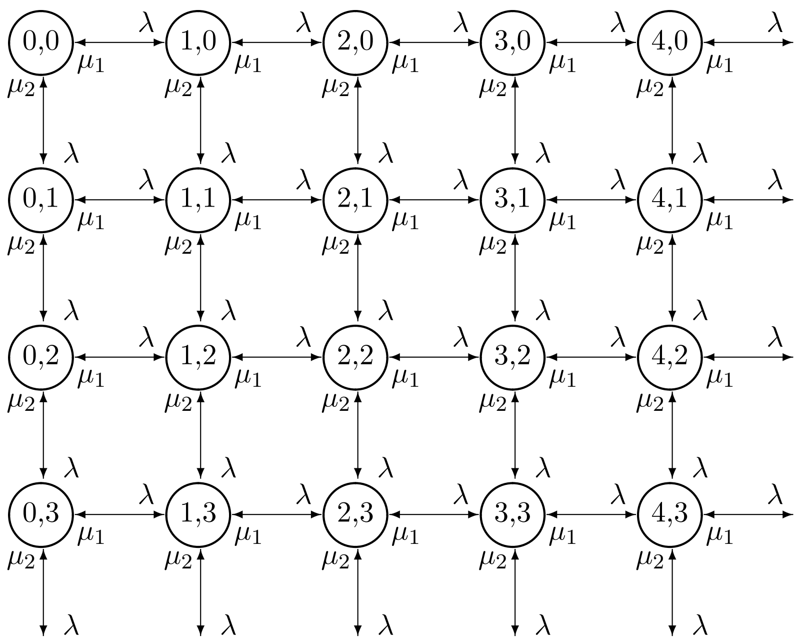


Figure 6.7: Reduced state diagram of two cascaded queues.

the network after service, then again the corresponding dropping of states and fluxes does not alter the balance of the fluxes of the infinite chain, and the solution is the same but the normalization constant (in this case the flow entering the second queue is still Poisson).

A third case is when the number in the entire network is limited, i.e., $N_1 + N_2 \leq m$. In this case, again, the corresponding dropping of states and fluxes does not alter the balance of the fluxes of the infinite chain, and the solution is the same but the normalization constant. This concept will be used in the next section.

6.2.4 Closed Cascaded Queues

Two cascaded queues.

Let consider the two cascaded queues of Section 6.2.2 and close the network assuming that the output of queue 2 is entered to the first one. The state space the pair (n_1, n_2) is redundant since $n_1 + n_2 = M$. Then we refer to the content n of one of them, say queue two. The state diagram is then the one of the Birth and Death process with upward and downward rate equal to μ_1 and μ_2 respectively. The solution is

$$\pi_i = \pi_0 \left(\frac{\mu_1}{\mu_2}\right)^i, \qquad i = 0, 1, \ldots, M. \tag{6.37}$$

The flow rate can be written

$$\lambda = \lambda_1 = \lambda_2 = \sum_{i=1}^{M} \pi_i \mu_2 = \mu_2(1 - \pi_0). \tag{6.38}$$

We can easily verify that (6.37) can be written as

$$\pi_{ij} = B\left(\frac{\lambda}{\mu_1}\right)^i\left(\frac{\lambda}{\mu_2}\right)^j, \qquad i+j=M, \quad i=0,1,\ldots,M, \tag{6.39}$$

where B is the normalization constant. The representation in (6.39) is rather interesting, since it shows that the joint distribution is the product of terms, such as $(\lambda/\mu_1)^i$, that represents, up to a constant, the distributions of the two queues when each of them is considered alone with Poisson arrivals of rate λ, i.e., exactly as with happens with the open queues of the previous section. However, here the arrivals at each queue are not Poisson, and the marginal distribution of the second queue is (6.37), different from $(\lambda/\mu_1)^i$, showing that the two queues are not statistical independent.

Since our example deals with a closed network, the rate λ can not be determined a priori, and, therefore, (6.39) seems useless. However, we know that λ_i is proportional to ν_i, the distribution of the routing chain. Therefore (6.39) can be rewritten as

$$\pi_{ij} = B'\left(\frac{\nu_1}{\mu_1}\right)^i\left(\frac{\nu_2}{\mu_2}\right)^j, \qquad i+j=M, \quad i=0,1,\ldots,M, \tag{6.40}$$

where in our case we have $\nu_1=\nu_2=1/2$, and the proportionality constant is included in B'.

We remark that the case of two queues where the first presents an infinite number of servers, each one of them at rate ν, the system becomes similar to the one we have already studied in Section 3.3.2(M/M/1 with finite population). The first queue, where each user has its proper server, in fact, represents the idle time of each users before entering the real service μ. Note also that the latter system becomes an open network of queues as M, and the number of servers in the first queue, become infinite, and ν decreases as $1/M$.

Three cascaded queues.
We add the queues of the preceding case a third queueing system (queue 3) with a single server of rate μ_3, whose output represent the input of the first queue, so that the three queues are closed in a circle and users migrates according to that circle. We also assume there are M users in the whole system. In this case the state of the whole network is represented by the pair (n_1, n_2), i.e., the occupancy of the first and the second queue (the occupancy of the third queue is $M-n_1-n_2$. The state diagram is given in Figure 6.8, where we have assumed $M=4$. It is exactly the diagram of Figure 6.6 where λ is replaced by μ, and the states where $n_1+n_2>4$ are dropped, together with the corresponding arrows in and out.

The dropped fluxes that enter and exit a single state are balanced; this means that the dropping of these fluxes does not alter the balance of fluxes with respect to the case in Figure 6.6. Therefore the solution is the same, up to a constant and the due substitutions, and we have

$$\pi_{ij} = \pi_{00}\left(\frac{\mu_3}{\mu_1}\right)^i\left(\frac{\mu_3}{\mu_2}\right)^j \tag{6.41}$$

Taking the marginal distributions we have

$$\pi_i^{(1)} = \sum_{j=0}^{M-i}\pi_{00}\left(\frac{\mu_3}{\mu_1}\right)^i\left(\frac{\mu_3}{\mu_2}\right)^j = \pi_{00}\left(\frac{\mu_3}{\mu_1}\right)^i f(i)$$

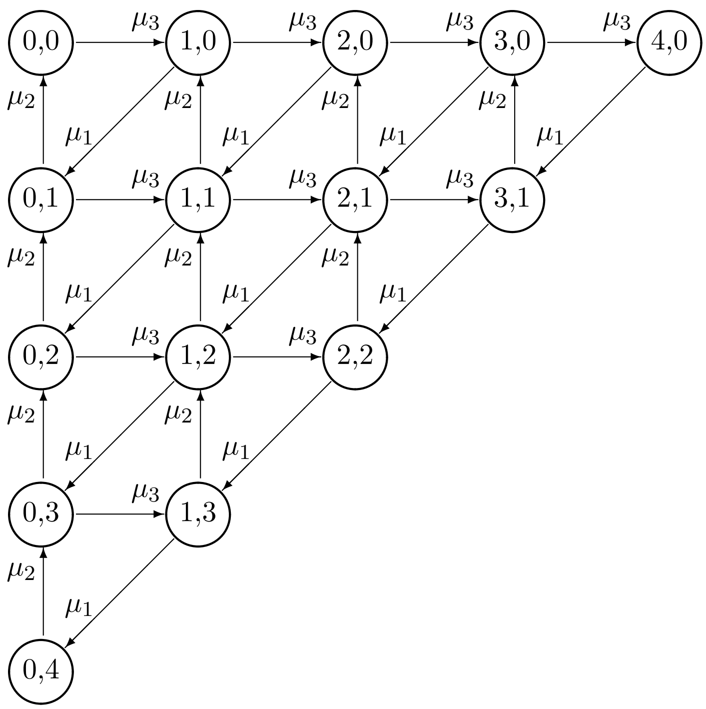


Figure 6.8: State diagram of the closed network of three queues

$$\pi_j^{(2)} = \sum_{i=0}^{M-j} \pi_{00} \left(\frac{\mu_3}{\mu_1}\right)^i \left(\frac{\mu_3}{\mu_2}\right)^j = \pi_{00} \left(\frac{\mu_3}{\mu_1}\right)^j g(j)$$

and, therefore, we have

$$\pi_{ij} \neq \pi_i^{(1)} \pi_j^{(2)},$$

i.e., the distributions are no longer independent.

Again (6.41) can be written as

$$\pi_{ijk} = B \left(\frac{\lambda}{\mu_1}\right)^i \left(\frac{\lambda}{\mu_2}\right)^j \left(\frac{\lambda}{\mu_2}\right)^k, \qquad i+j+k = M, \quad i = 0, 1, \ldots, M. \tag{6.42}$$

or, more exactly,

$$\pi_{ijk} = B' \left(\frac{\nu_1}{\mu_1}\right)^i \left(\frac{\nu_2}{\mu_2}\right)^j \left(\frac{\nu_3}{\mu_2}\right)^k, \qquad i+j+k = M, \quad i = 0, 1, \ldots, M, \tag{6.43}$$

with $\nu_1 = \nu_2 = \nu_3$.

Again, we see that the joint distribution of the three queues is the product, up to a constant, of the distributions of the three queues when each of them is considered alone with Poisson arrivals of rate λ. However, each term do not represent the marginal distribution and, therefore, the three queues are not statistical independent.

The results seen in this section, for both open and closed networks, generalize for any number queues, single server and multi server, and a specific generalization is exposed in Section 6.3.1.

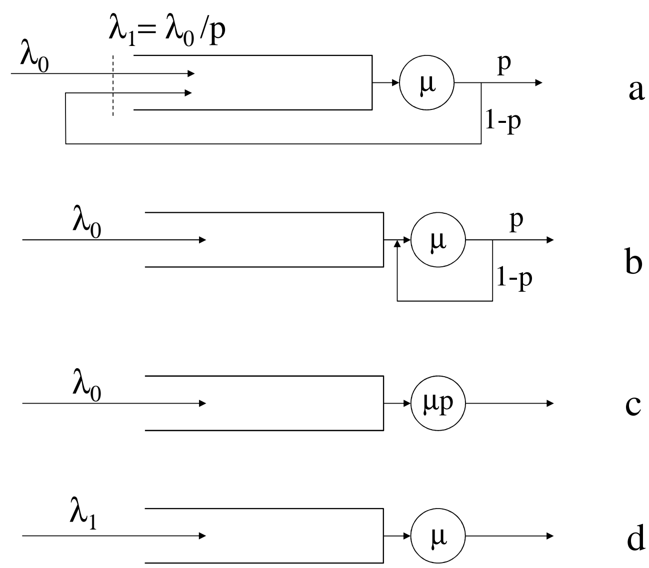


Figure 6.9: *Queues with service repetition.*

6.2.5 Closed paths

Property 6.33 excludes closed paths in order to have Poisson arrivals at each queue in an open markovian network. On the other side, in the closed network we have analyzed in the preceding section 6.2.4 there are closed paths (in a closed network all paths are closed), and we have seen the surprising results of (6.39) and (6.42), where each term represents the distribution of the corresponding queue with Poisson arrivals. This suggest that the joint distribution (and the marginals) in the open network can still be the same even in presence of closed paths. This can be verified directly if we introduce closed paths in the example in 6.2.2. Here, we show an example with a single queue.

Example (6.44)
Let consider a modified system M/M/1 where at end of the service the user decides to leave with probability p, while with probability $1-p$ *it returns to the queue with a new service time X independent of the previous one (Figure 6.9a).*

It is easily recognized that the occupancy is still a Markovian Birth and Death process with $\lambda_i = \lambda_0$ e $\mu_i = \mu p$, and

$$\pi_i = (1-\rho)\rho^i, \qquad \rho = \lambda_0/(\mu p).$$

The system behaves as the one in (Figure 6.9b), because the next service time is still taken independently of the position in the queue.

The same result can be obtained by observing that actually the sum of independent negative exponentials with rate μ in random number, having the number a Geometric distribution with probability success p, is still negative exponential with rate μp (Figure 6.9c).

The system proposed can also be seen as an open network of queues, in the special case of one queue with feedback. Evaluating the global flow in the queue we have

$$\lambda_1 = \lambda_0 + \lambda_1(1-p),$$

that yields

$$\lambda_1 = \frac{\lambda_0}{p},$$

and therefore the solution could be seen as that of a system with Poisson arrivals at rate λ_1 and service rate μ (Figure 6.9d). However,

Note (6.45)
the flow with rate $\lambda_1 = \lambda_0/p$ is no longer Poisson.

In fact, a return in $[t, t+\Delta t]$ depends on the entire arriving flow in $[-\infty, t]$; that is the returning flow depends on its past history. Note also that the leaving flow is still Poisson, as can be seen by Figure 6.9c.

We can also consider a similar closedpath returning, from the output of two cascaded queues to the first, with probability $1-p$, an extension of the case just seen. The state diagram is similar to the on in Figure 6.6 with the added probability flow from state (i, j) to stare $(i+1, j-1)$ with rate $\mu_2(1-p)$. The solution is easily verified to be the same as the original in Figure 6.6, by substituting λ with $\lambda_1 = \lambda/p$. Again, the flow entering queue 1, albeit not Poisson, provides a solution that is the same as the flow were Poisson.

6.3 Markovian Networks

The general approach to networks is to study the joint occupancy process at all nodes

$$\mathbf{n} = (n_1, n_2, \ldots, n_j, \ldots, n_J),$$

where n_j represents the content of node j (we explicitly warn the reader about the change of notation forced by the complexity of the new state space), and J is the number of nodes.

Markovian Networks require Poisson births, markovian routing with no restrictions (closed paths do not alter the markovian property of the network), negative exponential servers; this makes transitions at time t only depend on the state $\mathbf{n}$ at time t and not on the past history. We note that constraints are reduced with respect to Property 6.33. Here we further limit ourselves to networks where transitions at time t are only due to migrations of one user at a time. A migration of a user from queue j to queue k changes state $\mathbf{n}$ into state

$$T_{jk}\mathbf{n} = (n_1, n_2, \ldots, n_j - 1, \ldots, n_k + 1, \ldots, n_J),$$

where T_{jk} is the transition operator. Similarly, operator T_{j0} denotes a departure from the net at node j, while operator T_{0k} denotes an external arrival at node k. T_{00} is not considered. The balance equations are

$$\pi(\mathbf{n})\sum_{j=0}^{J}\sum_{k=0}^{J} q(\mathbf{n}, T_{jk}\mathbf{n}) = \sum_{j=0}^{J}\sum_{k=0}^{J} \pi(T_{jk}\mathbf{n})q(T_{jk}\mathbf{n}, \mathbf{n}), \qquad j \neq k. \tag{6.46}$$

Analytical solutions are only possible in special cases. Among these there is the case where the stationary distribution is a product form. It is therefore very important to understand whether there are conditions to say the solution is a product form. Among these we find networks called *Jackson's Networks.*

6.3.1 Jackson's Networks

Jackson's networks are markovian networks of both types, open and closed, where markovian servers can be in any number; however we have on the routing on transition rates, that must be of this type

$$\begin{aligned} &q(\mathbf{n}, T_{jk}\mathbf{n}) = \ell_{jk}\mu_j(n_j), \\ &q(\mathbf{n}, T_{j0}\mathbf{n}) = \ell_{j0}\mu_j(n_j), \\ &q(\mathbf{n}, T_{0k}\mathbf{n}) = \ell_{0k}\lambda, \end{aligned} \tag{6.47}$$

that is:

1. births to the network are of Poisson type;
2. the routing of a customer who comes from a generic queue j can not depend on distribution of customers in the queues, and is therefore described by the single matrix ℓ_{jk}; this rules out finite queues with blocking, since the blocked users can not be re-routed, including leaving the network; however closed paths are admitted;
3. the rate of exit from the generic queue j, when the queue is not empty, μ_j, may depend at most on the number of customers in the queue ($\mu_j(n_j)$), but not on the occupancy of the other nodes (e.g. blocked users can not be delayed), and not on k).

6.3.2 Open Jackson's Networks

Theorem: (6.48)

The asymptotic distribution of $\mathbf{N}(t)$ *for an open Jackson's network is a product form of the type*

$$\pi(\mathbf{n}) = \pi_1(n_1)\pi_2(n_2)\ldots\pi_J(n_J),$$

$$\pi_j(n_j) = \pi_j(0)\frac{\lambda_j^{n_j}}{\mu_j(1)\mu_j(2)\ldots\mu_j(n_j)} = \pi_j(0)\frac{\lambda_j^{n_j}}{\prod_{r=1}^{n_j}\mu_j(r)}, \tag{6.49}$$

where $\pi_j(n_j)$ is the marginal distribution at node j, and λ_j is the flow rate at node j.

Proof
We show first that λ_j is the flow rate at node j. In fact, it equals the departure rate, which is

$$\sum_{n_j>0} \pi(n_j)\mu_j(n_j) = \lambda_j \sum_{n_j>0} \pi(n_j-1) = \lambda_j,$$

where we have used (6.49). The last equality comes from the fact that the summation represents the summation of $\pi_j(n_j)$ over the entire state space.

We now show that (6.49) satisfy the balance equations. In fact, we easily see that such distribution satisfies the following more strict balance relation

$$\pi(\mathbf{n}) \sum_{k=0}^{J} q(\mathbf{n}, T_{jk}\mathbf{n}) = \sum_{k=0}^{J} \pi(T_{jk}\mathbf{n}) q(T_{jk}\mathbf{n}, \mathbf{n}), \tag{6.50}$$

which, in turn, satisfies (6.46). We note that relation (6.50) imposes that the flows due to a departure from node j is balanced by arrivals and births at the same queue, including node 0. To prove that (6.50) is indeed verified by (6.49), we rewrite it insulating the flows to and from node 0 as

$$\pi(\mathbf{n})q(\mathbf{n}, T_{j0}\mathbf{n}) + \pi(\mathbf{n}) \sum_{k=1}^{J} q(\mathbf{n}, T_{jk}\mathbf{n}) = \pi(T_{j0}\mathbf{n})q(T_{j0}\mathbf{n}, \mathbf{n}) + \sum_{k=1}^{J} \pi(T_{jk}\mathbf{n})q(T_{jk}\mathbf{n}, \mathbf{n}), \tag{6.51}$$

$$\pi(\mathbf{n}) \sum_{k=1}^{J} q(\mathbf{n}, T_{0k}\mathbf{n}) = \sum_{k=1}^{J} \pi(T_{0k}\mathbf{n})q(T_{0k}\mathbf{n}, \mathbf{n}). \tag{6.52}$$

Using (6.47), and the following substitutions

$$\begin{aligned} q(T_{jk}\mathbf{n}, \mathbf{n},) &= \ell_{kj}\mu_k(n_k+1), \\ q(T_{0k}\mathbf{n}, \mathbf{n},) &= \ell_{k0}\mu_k(n_k+1), \\ q(T_{j0}\mathbf{n}, \mathbf{n},) &= \ell_{0j}\lambda_0, \end{aligned} \tag{6.53}$$

together with, from (6.49),

$$\pi(T_{jk}\mathbf{n}) = \pi(\mathbf{n}) \frac{\mu_j(n_j)}{\lambda_j} \frac{\lambda_k}{\mu_k(n_k+1)}, \qquad j,k \neq 0, \tag{6.54}$$

$$\pi(T_{j0}\mathbf{n}) = \pi(\mathbf{n}) \frac{\mu_j(n_j)}{\lambda_j}, \qquad j>0, \tag{6.55}$$

$$\pi(T_{0k}\mathbf{n}) = \pi(\mathbf{n}) \frac{\lambda_k}{\mu_k(n_k+1)}, \qquad k>0, \tag{6.56}$$

into(6.51) and simplifying we get

$$\lambda_j \sum_{k=0}^{J} \ell_{jk} = \sum_{k=0}^{J} \ell_{kj}\lambda_k,$$

where both terms indeed represents λ_j, while substituting into(6.52) we get $\lambda_0 = \lambda_0$. ♣

Theorem 6.48 shows that the occupancies $N_j(t)$ are statistically independent RV, and that the marginal distributions are those we have when the queue is taken alone with a Poisson arrival process of rate λ_j. In fact, expression (6.49) is the distribution of the content of such queues when they are represented by a Birth and Death process. In particular, if queue j is for a single server queue, denoted $\rho_i = \lambda_i/\mu_i$, we have

$$\pi_i(k) = (1-\rho_i)\rho_i^k. \tag{6.57}$$

Example (6.58)
Given an open network of three cascaded single-markovian-server queues, 1, 2 and 3, with service rates μ_1, μ_2, e μ_3. Poisson users enter the network at rate λ and all reach queue 1. After leaving queue i users enter $i+1$ queue with probability p_i, whereas with probability $1-p_i$ return to queue 1, except for queue 3, where with probability p_3 the users leave the network. Find the load factor of each server and the upper limit of λ for which the system is stable, in the cases $\mu_1 = 2$, $\mu_2 = 3$, e $\mu_3 = 1$, $p_1 = 0.8$, $p_2 = 0.5$, $p_3 = 0.3$.

Thanks to Theorem 6.19 the flow rates at queues are given by

$$\lambda_1 = \frac{100}{12}\lambda, \qquad \lambda_2 = \frac{100}{15}\lambda, \qquad \lambda_3 = \frac{100}{30}\lambda,$$

and the load factors

$$\rho_1 = \frac{100}{24}\lambda, \qquad \rho_2 = \frac{100}{45}\lambda, \qquad \rho_3 = \frac{100}{30}\lambda.$$

Hence, we see that the most loaded queue is 1 and, for stability, we must have $\lambda < 24/100$.♣

We also cite, without proving, the following properties:

Theorem: (6.59)
If it is possible to decompose the open network R into many disjoint sub-networks R_i such that we can move from a sub-network to the next without being able to return to any of the previous sub-network, then the user flows from one sub-network to another are Poisson.

It turns out that flows other than the ones specified above are neither Poisson nor even renewal events.

We remark again that the previous theorems extends the Property 6.33, that a single queue behaves as if it was isolated with Poisson arrivals, although the actual arrival process is far from being Poisson.

Theorem: (6.60)
The PASTA property also extends to Open Jackson's networks.

The above theorem allows to evaluate delays, as the distribution seen by arrivals coincides with time continuous one. Since the latter is the one we have with the queue in isolation, also the delay distribution and moments coincide with the ones we have with the queue in isolation.

6.3.3 Blocking systems

The distribution seen in Jackson's Networks also extends, in some cases, to blocking systems, i.e., systems that have finite capacity, e.g., system k is such that $0 \leq n_k \leq Q_j$. This is possible under the condition that blocked users do not alter their routing mechanism, otherwise we would have a routing mechanism, $\ell_{jk}(n_k)$, that would depend on n_k, which is not allowed by Jackson's Networks. However, the routing mechanism is not altered, i.e., does not depend on n_k, if, upon being blocked, they immediately reach the output of the system. In fact, this effect can be taken into account in Jackson's Networks by assuming

$$\mu_j(n_j) = \infty, \qquad n_j > Q_j.$$

With condition above, user enters the system even if $n_j > Q_j$, but, in this case, it, or the user in service, immediately reaches the output because the service rate is infinite (since the service time is negative exponential there is no need to say which user reaches the output).

For example, distribution (6.49) still holds with the admitted range of n_j, since it is zero for $n_j > Q_j$, and $\pi_j(0)$ must be re-computed.

What said above matches with the observation done in Chapter 3 that the output flow of such a blocking queue is still Poisson.

6.3.4 Closed Jackson's Networks

In closed networks the number of users is the constant M. Therefore, the state space of $\mathbf{n}$ is such that

$$\sum_j n_j = M. \tag{6.61}$$

If we go back to the proof of (6.49) for open networks, we see that, with constraint (6.61) balance equation (6.52) does not exits, while also in (6.51) the terms outside the summation are canceled. However, even with these changes, (6.49) is the solution, up to constant and also represents the distribution of the closed network.

However, in this case rates λ_j can not be derived independently. Fortunately, we know that rates λ_i are always proportional to probabilities ν_i of the routing chain (Theorem 6.28). So we can write

Theorem: (6.62)

$$\pi(\mathbf{n}) = B_M \psi_1(n_1)\psi_2(n_2)\ldots\psi_J(n_J),$$

$$\psi_j(n_j) = \frac{\nu_j^{n_j}}{\prod_{r=1}^{n_j} \mu_j(r)}. \tag{6.63}$$

where B_M is the new normalization constant.

Unfortunately, the marginal distributions, for which there is not a general closed form expression, are such that there is not the independence there is in open networks. So, many results of open networks are difficult to attain. Some results are attained in implicit form. The proportional constant between flow rates and distribution ν_i is provided by the following;

Theorem: (6.64)

The flow rate at queue j is

$$\lambda_i = \nu_i \frac{B_M}{B_{M-1}}, \tag{6.65}$$

where B_M is the normalization constant with M users.

The proof comes starting from the expression

$$\lambda_i = \sum_{\Omega} \mu_i(n_i)\pi(\mathbf{n}) = \sum_{\Omega} \mu_i(n_i) B_M \prod_{j=1}^{J} \frac{\nu_j^{n_j}}{\prod_{r=1}^{n_j} \mu_j(r)}$$

which can be written as

$$\lambda_i = \frac{B_M \nu_i}{B_{M-1}} \sum_{\Omega} \left[B_{M-1} \frac{\nu_i^{n_i-1}}{\prod_{r=1}^{n_i-1} \mu_i(r)} \prod_{j=1}^{\substack{j\neq i\\ J}} \frac{\nu_j^{n_j}}{\prod_{r=1}^{n_j} \mu_j(r)} \right] = \frac{\nu_i B_M}{B_{M-1}}$$

The last equality holds because the summation of terms inside the parentheses is one, since it represents the joint distribution of the system with $M-1$ users.

By Theorem 6.64 we see that, as expected, as M changes also rates λ_i change. Also (6.65) still holds if, there and in the solution (6.63), we substitute ν_i with $k\nu_i$, modifying the normalizing constant into $B'_M = B_M/k^M$, since k simplifies.

We have also the following modification to the PASTA property

Theorem: (6.66)

The users that enter the queue j, see a number of users already there whose distribution is the marginal obtained from (6.63), when the number users in the network is $M-1$.

This is an interesting result, whose explanation lies in the fact that an arriving users can never see himself already in the queue.

6.3.5 Cascaded M/M/1 queues

Here we consider a closed network of J cascaded queues with a single server. The solution of the routing matrix ν_i is uniform, owing to the symmetry; therefore it can be considered in constant B_M in (6.63), which can be written as

$$\pi(n_1, n_2, \ldots, n_J) = B_M \left(\frac{1}{\mu_1}\right)^{n_1} \left(\frac{1}{\mu_2}\right)^{n_2} \cdots \left(\frac{1}{\mu_J}\right)^{n_J}, \tag{6.67}$$

The evaluation of the constant B_M may be very complex because the state space is very large even for small number of queues J and M. The size of the space can be computed by noting that each state can be represented by $M+J-1$ consecutive positions among which we choose $J-1$ queue boundaries, while the remaining M represent the users. This choice can be made in

$$\binom{M+J-1}{J-1} \tag{6.68}$$

ways, which represents therefore the number of possible states.

If the rates $\mu_i = \mu$ are all equal, distribution (6.67), becomes

$$\pi(n_1, n_2, \ldots, n_J) = B_M \mu^{-M} = \binom{M+J-1}{J-1}^{-1}. \tag{6.69}$$

The flow rate, by (6.65) is

$$\lambda = \frac{M}{M+J-1}\mu,$$

and the load factor

$$\rho = \frac{M}{M+J-1}.$$

Starting from the joint distribution we can prove that

$$P(N \geq k) = \frac{B_M}{B_{M-k}} = \frac{M}{M+J-1}\,\frac{M-1}{M+J-2}\cdots\frac{M-k+1}{M+J-k},$$

where N is the content of the generic queue. This provides the marginal distribution as

$$P(N = k) = P(N \geq k) - P(N \geq k+1), \tag{6.70}$$

and shows how complex is the distribution even in this simple case. We can also prove that

$$E[N] = \sum_{k=1}^{M} P(N \geq k) = \binom{M+J-1}{J-1}^{-1} \sum_{k=1}^{M} \binom{M-k+J-1}{J-1}$$

which turns out to be:

$$E[N] = \frac{M}{J}, \tag{6.71}$$

as obvious. We have also

$$E[V] = \frac{E[N]}{\lambda} = \frac{M+J-1}{J} 1/\mu. \tag{6.72}$$

Note that with an open network with the same load ρ we have

$$E[V'] = \frac{1}{1-\rho} 1/\mu = \frac{M+J-1}{J-1} 1/\mu,$$

greater than in the closed network.

6.3.6 Mean value analysis

In the previous section we have seen that occupancy parameters such as the flow rate, the average occupancy and so on, are difficult to attain in a closed networks. Even constant B_M is hard to attain even in networks where M and J are moderately large, because the number of elements of the distribution increases exponentially.

Here we show a way to get these parameters whose complexity is moderate.

We refer here to the case where each queueing system has only one server, but the method can be extended to more servers.

Referring to queueing system j and to a number M in the network, we denote:

$n_j(M) = E[N_j(t)]$ the average number of customers in the system (service included);

$\tau_j(M) =$ the mean sojourn time in the system (service included);

$\lambda_j(M) =$ the arrival rate;

Thanks to Theorem 6.66 and to the memoryless property of the negative exponential, we have

$$\tau_j(M) = \frac{1}{\mu_j} + \frac{1}{\mu_j} n_j(M-1). \tag{6.73}$$

Referring to Theorem 6.28 we have

$$\lambda_j(M) = \frac{\nu_j M}{\sum_{i=1}^{J} \nu_i \tau_i(M)}. \tag{6.74}$$

Finally, by the Little's result:

$$n_j(M) = \lambda_j(M)\tau_j(M). \tag{6.75}$$

Relations (6.73), (6.74) e (6.75), can be recursively solved in M, starting from

$$n_k(0) = 0.$$

6.3.7 Blocking with sessions

A user's session is defined as an alternation of calls and idle periods, with a call at the beginning and at the end of the session. The system is markovian if we assume that calls are negative exponential distributed with parameter μ, that idle periods are negative exponential distributed with parameter ν, and that at end of each call a new call is attempted, after an idle period, with probability $\alpha > 0$; which means that the number of calls in a session are geometrically distributed with average $1/(1-\alpha)$.

We assume that each user upon a call enters an $M/M/m$ blocking system, and at the end of the call reaches, with probability α, an $M/M/\infty$, where each service time represents the idle period of the session. We also assume that when a call is blocked, the user reaches, with probability α, the $M/M/\infty$ system. Then the whole system is a Jackson's Network, where the two systems can be considered separately. In particular, the blocking system presents an input traffic with rate $\lambda/(1-\alpha)$, and the blocking probability is still provided by the the Erlang-B formula we derived in Chapter 5.

6.4 General Routing

Property 6.33, assures that in an open network of queues without closed paths, with Poisson births and negative exponential servers, and with markovian routing, all the queues behave as if they were M/M systems in isolation.

It is easy to see that this property still holds when we have many different users' types, each one originated by its own Poisson flow and having its markovian routing matrix $\ell_{ik}^{(r)}$. In fact, the superposition of Poisson flows, mix different types at random, and the probabilistic mixing is not altered by service disciplines like FCFS, LCFS or RO. The mixing is altered by routing, but, again it remains random so that collective flows in the network are still Poisson. This occurs in spite of the fact that the global routing process, with matrix

$$\ell_{ik} = \sum \alpha_r \ell_{ik}^{(r)}. \tag{6.76}$$

is no longer markovian.

When we have closed paths and user flows are no longer Poisson, then Property 6.33 no longer holds. As an example, let consider one single-server queue and exponential service with rate μ where users enters the networ/queue as a Poisson flow of rate λ. Served users re-enter the the queue only once at end of the line. In order to know whether the user under service is there for the first or the second time, we must memorize in the state the entire positions of the users, e.g. $\mathbf{s} = (s_1, s_2, \ldots, s_n)$, the index being the position of the users in the system, position $n = 1$ the server, n the number of users in the system, and s_i their *stage*, 1 or 2. This makes the state space rather complex; however, transition are, in this case, rather simple. For example, one particular balance equation is

$$\pi(s_n = 2)(\lambda + \mu) = \pi(s_1 = 1, s_n = x)\mu + \pi(s_1 = 2, s_{n+1} = 2)\mu,$$

which is satisfied by the uniform distribution

$$\pi(s_1, s_2, \ldots, s_n) = (\lambda/\mu)^n. \tag{6.77}$$

It is easy to show that all other possible transitions are satisfied by the same distribution, which is then the asymptotic distribution. Furthermore, the number od states with n users in the system is 2^n, so that the distribution of the number in the system, i.e., the probability we have n users, is

$$\pi_n = \pi_0 \left(\frac{2\lambda}{\mu}\right)^n, \qquad n \geq 0. \tag{6.78}$$

We note that the solution above is the one of an M/M/1 system with arrivals at rate 2λ. We also note that the distribution is completely different if users return to the queue in position 1. In this case, in fact, users do not leave service after the first service period and, therefore, the service time becomes an Erlang-2 variable, with average $2/\mu$, and a more complex π_n. The former system can also be seen as a system where the service time becomes an Erlang-2 variable but with an interruption after a the first stage of the Erlang-2 (a negative exponential variable), after which the user reaches the end of the queue. This explain the change in the distribution π_n. Also note that the result is exactly the one we have with the processor-sharing service discipline.

6.4.1 Kelly's networks

The approach shown above, and (6.77), can be extended to an entire network with multiple paths with memory. In order to take into account multiple, and random, paths, we must introduce *user types*, t, so that to each user type correspond a different path; other assumptions are as in Jackson's Networks, in particular different user types must have the same negative-exponential service-time pdf.

Denoted by t_{jl} the type of the user in position l in queue j, and similarly, s_{jl}, the *stage* reached on its path, we define the *class* of such user as

$$c_{jl} = (t_{jl}, s_{jl}).$$

Then, vector

$$\mathbf{c}_j = (c_{j1}, c_{j2}, ...c_{jn_j})$$

describes the state of the queue, and the vector of vectors

$$\mathbf{c} = (\mathbf{c}_1, \mathbf{c}_2, ...\mathbf{c}_J)$$

describes the state of the entire network.

Denoting by $\nu(t)$ the flow rate of users of type t, we denote by $\lambda_j(c)$ the rate of the flow entering queue j and due to users of class $c = (t, s)$, we have

$$\lambda_j = \sum_c \lambda_j(c). \tag{6.79}$$

Then, in open systems we have the following theorem whose proof is omitted,

Theorem: (6.80)

$$\pi(\mathbf{c}) = \pi_1(\mathbf{c}_1)\pi_2(\mathbf{c}_2)\ldots\pi_J(\mathbf{c}_J),$$

$$\pi_j(\mathbf{c}_j) = b_j \prod_{l=1}^{n_j} \frac{\lambda_j(c_{jl})}{\mu_j(l)}. \tag{6.81}$$

If we denote by $n(c_j)$ the number of users in system j with the same class c_j we can write

$$\pi_j(\mathbf{c}_j) = b_j \frac{\prod_{c_j} \lambda_j(c_j)^{n_j(c_j)}}{\prod_{l=1}^{n_j} \mu_j(l)}, \qquad \sum_{c_j} n_j(c_j) = n_j. \tag{6.82}$$

The number of class configurations that presents the same sequence $(n_j(1), n_j(2), \ldots)$, and thus the same (6.82), is

$$\frac{n_j!}{n_j(1)!\, n_j(2)! \ldots},$$

and the probability $\pi_j(n_j)$ we have n_j users in queue-system j is the summation of (6.82) over all $\mathbf{c}_j$ that sum to n_j:

$$\pi_j(n_j) = \frac{1}{\prod_{l=1}^{n_j} \mu_j(l)} \sum_{n_j(1), n_j(2), \ldots} \frac{n_j!}{n_j(1)!\, n_j(2)! \ldots} \prod_{c_j} \lambda_j(c_j)^{n_j(c_j)}, \qquad \sum_{c_j} n_j(c_j) = n_j. \tag{6.83}$$

The summation above coincides with the multinomial expansion

$$\sum_{n_j(1), n_j(2), \ldots} \frac{n_j!}{n_j(1)!\, n_j(2)! \ldots} \prod_{c_j} \lambda_j(c_j)^{n_j(c_j)} = \left(\sum_{c_j} \lambda(c_j) \right)^{n_j} = \lambda_j^{n_j},$$

and, therefore, we have proven the following

Corollary (6.84)

$$\pi(\mathbf{n}) = \pi_1(n_1)\pi_2(n_2) \ldots \pi_J(n_J),$$

$$\pi_j(n_j) = \pi_j(0) \frac{\lambda_j^{n_j}}{\prod_{l=1}^{n_j} \mu_j(l)}. \tag{6.85}$$

The Corollary just proven shows that the solution that holds for the open Jackson's networks, also holds for general routing.

It can be proved that also in this case holds the ASTA property, that is, the distribution upon arrivals at a queue is the same as the continuous-time one, namely (6.81) and (6.85).

6.5 Open Networks of Symmetric Queues

The results seen in the preceding sections are striking in their simplicity. However, they all require that users have the same service-time pdf, which must be negative exponential. Those results can be extended to a general service-time pdf for a class of queues called *symmetric*, which require a special class of service disciplines.

Let define a more general framework for service disciplines. In order to deal with servers with different service speed we must decouple the reqiored service, χ, from the service speed ϕ. We require

a) the service requirement is RV χ with average m_χ;

b) the overall service rate with i users is $\phi(i) > 0$;

c) a fraction $\gamma(l,i), \sum_l \gamma(l,i) = 1$, of the service speed is given to the users u_l in position l in the queue, $1 \leq l \leq i$; when this user leaves service users in positions $l+1, l+2, \ldots, i$, move in positions $l, l+1, \ldots, i-1$;

d) when a user arrives at the queue with i users, it enters the system in position l, $l = 1, \ldots, i+1$ with probability $\delta(l, i+1), \sum_l \delta(l,i) = 1$; users previously in positions $l, l+1, \ldots, i$, move in positions $l+1, l+2, \ldots, i+1$; respectively;

This framework is able to represents many service disciplines. For example

i - FCFS M/G/1 by setting $\phi(i) = \phi$, $\gamma(1,i) = 1$, $\gamma(l,i) = 0, l \neq 1$, $\delta(i,i) = 1$, $\delta(l,i) = 0, l \neq i$; the average service time is $m_X = m_\chi/\phi$.

ii - LCFS M/G/1 with preemption by setting $\phi(i) = \phi$, $\gamma(1,i) = 1$, $\gamma(l,i) = 0, l \neq 1$, $\delta(1,i) = 1$, $\delta(l,i) = 0, l \neq 1$;

iii - Random Order M/G/1 with preemption by setting $\phi(i) = \phi$, $\gamma(1,i) = 1$, $\gamma(l,i) = 0, l \neq 1$, $\delta(l,i) = 1/i, \quad \forall l$;

iv - M/G Processor sharing by setting $\phi(i) = \phi$, $\gamma(1,i) = 1/i, \forall l$;

v - M/G/∞ by setting $\phi(i) = i\phi$ e $\gamma(l,i) = 1/i$; each user has its own server at rate ϕ.

vi - M/G/m/m by setting

$i \leq m$ - $\phi(i) = i\phi$ e $\gamma(l,i) = 1/i$; each user has its own server at rate ϕ.

$i > m$ - $\phi(i) = \infty$, $\gamma(l,i) = 0, l < m$, $\gamma(l,i) = \frac{1}{i-m}, m < l \leq i$; this deliver an infinite service rate to users that upon arrival find m users in the system; this causes these users to immediately leave the system. This account for the fact that such users are, in the real system blocked and cleared.

Symmetric queues are attained by the preceding framework by setting $\delta(l,i) = \gamma(l,i)$ (hence the term symmetric). Unfortunately, this constraint can not model such important system as FCFS.

With symmetric queues the service discipline is such that all users in the system have already received some amount of service. In fact, any new arrival changes the service speed of the others, and then is an preemptive one.

For a network of symmetric queues and general service time and general routing we have

Theorem: (6.86)

The asymptotic distribution of $\mathbf{N}(t)$ is a product form of the type

$$\pi(\mathbf{n}) = \pi_1(n_1)\pi_2(n_2)\ldots\pi_J(n_J),$$

$$\pi_j(n_j) = \pi_j(0)\frac{(\lambda_j m_j)^{n_j}}{\prod_{l=1}^{n_j}\phi(l)}, \tag{6.87}$$

where $\pi_j(n_j)$ is the marginal distribution at node j, λ_j is the flow rate at node j, and m_j is the average service requirement at node j.

We note that, again, (6.87) represents the distribution of a Birth and Death process. When we have one constant-rate server(6.87) becomes

$$\pi_j(n) = (1-\rho)\rho^n \qquad n \geq 0 \tag{6.88}$$

with $\rho = \lambda m_\chi/\phi$, which is the geometric distribution we get with an M/M/1. This confirm what we reported in Chapter 3 about Processor Sharing Discipline.

For a blocking system with m servers and no queue the (6.87) becomes

$$\pi_i = \pi_0 \frac{(m\rho)^i}{n!} \qquad i \leq m \tag{6.89}$$

i.e., the Erlang distribution. In this case the symmetric queue requirement has no effect, so that this result is a proof that it holds also for general service, as mentioned in Chapter 3.

Again, we have the extension of the PASTA property to the arrivals at any queue, which allows to evaluate delays.

6.6 Problems for solution

P.6.1 (4.20) A Poisson flow at rate λ enters node 1 and subsequently enters node 2. At the exit of node 2 users go back to node 1 with probability p, otherwise they leave the network. A second Poisson flow at rate γ enters node 2 and, afterwards goes to node 1 with probability q, otherwise leaves the network. The users of this flow that exit node 1 enter node 2 with probability q, q, otherwise they leave the network. Assuming that in all nodes users spend an average time given by

$$V(\Gamma_i) = \frac{1}{1-\Gamma_i}$$

where Γ_i is the global flow rate at node i, find

a) the global rate Γ_i of the flow at each node i;

b) the average number of users of each type at each node.

P.6.2 (4.26) In a closed network of 5 nodes i, $i = 1, 2, 3, 4, 5$ users remain in each node an average time equal to the index of the node. The routing matrix is such that from node i they reach node 1 with probability $p = 1/2$, and with probability $1/2$ they reach the next node in sequence (or node 1 when $i = 5$). Assuming there are $M = 10$ customers in the system, find

- the flow rate at each node;
- the asymptotic probability that a user is in node i at time t.

P.6.3 (7.4) An open network of J markovian, single-server queues with service rate μ is connected in a circle. Users enter the network at queue 1, according to a Poisson flow of rate λ_0, and leave the system immediately before queue 1 with probability p. Find

a) the maximum λ_0 allowed by network stability and the server load factor;
b) the average network delay;
c) the average number of visits to each queue.

P.6.4 (7.5) An open network of 3 markovian, single-server queues with service rates μ_i, $(i = 1, 2, 3)$ are connected in a cascade. Users enter the network at queue 1, according to a Poisson flow of rate λ. At the output of queue i they reach queue 1 with probability $1 - p_i$, while with probability p_i that enter queue $i+1$, except for queue 3 where with probability p_i they leave the network. Assuming

$$\mu_i = i \quad p_1 = 0.8, \quad p_2 = 0.5, \quad p_3 = 0.3,$$

find

a) the flow rate at each queue;
b) the maximum λ allowed by network stability;
c) The average time in each queue at each entrance and in the entire life, assuming $\lambda = 0.01$;
d) the average network delay, again with $\lambda = 0.01$.

P.6.5 (7.12) Three markovian, single-server queues with rate $\mu = 1$ are closed in a circle. Assuming there are 3 users in the network, find the rate of flow, the load factor of each queue and the content distribution of queue 1.

P.6.6 (7.15) An open network of 3 markovian, single-server queues with service rate μ has the following routing matrix

$$\begin{vmatrix} 0 & 0.5 & 0 & 0.5 \\ 0 & 0.5 & 0.5 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0.25 & 0.25 & 0.5 & 0 \end{vmatrix}$$

where 0 is the source node and λ the source rate of the Poisson input flow. Find

a) The bottleneck queue, i.e., the one that first becomes unstable when λ increases;
b) the average network delay;
c) the average network delay of a user that enters the network at node 1;
d) the average number of time a user enters queue 1, and the same figure for users that enter the network at node 1.

P.6.7 (7.16) A closed network of 3 markovian, single-server cascaded queues with service rates $1, 2$, and 3 respectively has 2 users. Find the load factor and the occupancy distribution of queue 1.

P.6.8 (7.24) An open network of two queues with a single server at rate μ has two types of users, A and B. The first type enters the network according to a Poisson flow of rate λ_{0A}, reaches queue 1, then queue 2, then goes back to queue 1 with probability p, repeating the process, otherwise leaves the system. The second flow of B users enters queue 2 according to a Poisson flow of rate λ_{0B}, then enters queue 1 with probability q, otherwise leave the system. These users, after leaving queue 1 enter queue 2 with probability q, repeating the process, otherwise leave the system. Find

a) the flow rate at each queue and the maximum allowed by stability assuming $\lambda_{0A} < \lambda_{0B}$;

b) the average number of users of each type at each queue;

P.6.9 (7.22) In an open network of three queues,1,2 and 3, with single server at rates $\mu, 2\mu, \mu$, respectively a Poisson flow of rate λ enters node 1 and goes, with equal probability, toward nodes 2 and 3, and then leaves the network. Another Poisson flow of rate 2λ enters node 2 and goes, with equal probability, toward nodes 1 and 3, and then leaves the network. A third Poisson flow at rate λ enters at node 3. We have two ways to route it: A toward node 1 and then out of the network; B toward node 2 and then out of the network. Find

a) the global flow rate at nodes in the two cases A and B;

b) which routing strategy, between A and B, minimizes the network delay of the third flow.

c) which routing strategy, between A and B, minimizes the network delay of the third flow.

Evaluation can be done with $\lambda = 1$ and $\mu = 5$.

P.6.10 (7.23) A closed network is composed of three cascaded queues, each with a single markovian server at rate μ. Denoted by M the number of users in the network, find

a) The occupancy distribution π_i, $i = 0, 1, 2, 3$ of the first queue in the case $M = 3$;

b) The flow rate at the queues and the average time it takes for a customer to make a complete circle;

c) The average network delay of an open network with the same cascaded queues with a flow whose rate is equal to the one found in b).

P.6.11 In an open network of three single server queues of rate μ, users arrive at node 1 according to a Poisson flow of rate λ. Then, they are routed along the following path: $1 \to 2 \to 1 \to 3 \to 1 \to 3 \to 0$. Find the network delay.

P.6.12 (7.3) In an open network of two cascaded queues with a single server of rate μ, users arrive at queue 1 according to a Poisson process of rate λ. Users that leave queue 1 enter queue 2, and those leaving queue 2 enters queue 1 with probability p, otherwise leave the network. Draw the state diagram and show that the joint distribution is the product of the form of the distributions of the queues as if they were isolated.

P.6.13 (7.30) In an open network of two cascaded queues users arrive at queue 1 according to a Poisson process of rate λ. Users that leave queue 1 enter queue 2, and those leaving queue 2 enters queue 1 with probability $1-p$, otherwise leave the network. Queue 1 has a markovian single server of rate 2μ, while queue 2 has two markovian servers at rate μ and no waiting room and users that can not enter the queue leave the network.

a) Draw the state diagram (n_1, n_2) and check wheter the solution is the product form of the distributions of the queues in isolation.

b) Find the average network delay .

P.6.14 - A closed network of 4 markovian queues, 1,2,3,4, has $M = 100$ users and the following routing matrix

$$\begin{vmatrix} 0 & 0.5 & 0 & 0.5 \\ 0 & 0.5 & 0.5 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0 & 0 & 0.5 \end{vmatrix}$$

All the queues have *infinite* servers whose rates are equal to the index of the node. Find

a) the flow rates at each queue;
b) the average delay spent in the subnetwork composed of nodes $2, 3, 4$, by a user between the exit and the return to node 1;

(Note that with an infinite number of servers there is no queue and the time in the node is equal to...)

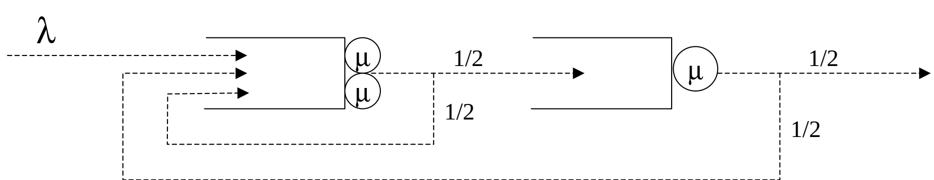
P.6.15 - An open network of 4 markovian queues, 1,2,3,4, serves two Poisson flows of users, the first at rate λ, enters the network at node 1, while the second at rate 2λ, enters the network at node 2. The two flows have different routing matrices among nodes 0,1,2,3,4, as given below

$$\begin{vmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \end{vmatrix} \qquad \begin{vmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0.5 & 0 & 0 & 0.5 \\ 1 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0.5 & 0 \end{vmatrix}$$

Queue 2 has an *infinite* number of markovian servers, each of rate μ, while queues 1,3 and 4, have a single server at rates 2μ. Find

a) the bottleneck queue, ie, the one with highest loading factor ρ;
b) the average network crossing delay;
c) the average network crossing delay of users of the first flow;

P.6.16 - The open network in the figure below is composed by two systems, one with two servers and the other with one server, all with service rate μ. The input flow is Poisson, with the routing probabilities shown in the figure. Find



a) the global (between birth and death) average time spent by a user in the system with one server.
b) the average number of users in the network.
c) the same as in a) and b) assuming that the input flow is dropped, so that the network becomes closed, and that there are three users in the network.

P.6.17 - An open network of 3 markovian, single-server queues with service rate μ has the following routing matrix

$$\begin{vmatrix} 0 & 0.4 & 0 & 0.6 \\ 0 & 0.6 & 0.4 & 0 \\ 0.5 & 0.25 & 0 & 0.25 \\ 0 & 0.5 & 0.5 & 0 \end{vmatrix}$$

where 0 is the source node and 2λ the source rate of the Poisson input flow. A second Poisson flow of rate $\lambda/2$ enters the network at node 2, then goes round to nodes 3, 1 and back to 2. This for exactly two rounds and finally it leaves the network. Find

a) The bottleneck queue, i.e., the one that first becomes unstable when λ increases, and the corresponding maximum λ;
b) the average network crossing delay, and the average network crossing delay suffered by the second flow.
c) the average of the total time spent by a user of the second flow in node 3.

P.6.18 - Compare the average crossing delay of the following three queuing system, with markovian servers at rate μ, when arrivals are Poisson at rate λ:

a) two cascaded single-server systems at rate μ;
b) one single-server system at rate μ, when after completing service users go back to the queue with probability 1/2, and otherwise leaves;
c) double server system, each server at rate $\mu/2$.

P.6.19 - A Poisson flow with rate λ enters a network from node 0 and is routed according matrix given below.

$$\begin{vmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \end{vmatrix}$$

Assuming that each user stays in each node for a time equal to T, find

a) the flow rate at each node;
b) the average number of users in the network of nodes 1,2,3,4, and its average crossing delay (between birth and death).

Answer again to questions a) and b) in the case where the number of users M in the entire network, including node 0, is M, and that the sojourn time in node 0 is M/λ (the network including node 0 is a closed network). What if $M \to \infty$?

P.6.20 Compare the average crossing delay of the following two queuing systems, with markovian servers, when arrivals are Poisson at rate λ:

a) two cascaded double-server systems at rate μ;
b) one single-server system at rate 2μ when users after service re-enter the system with probability 1/2.

P.6.21 A closed network of two nodes is such that users have a sojourn time equal to T and $2T$ in the two nodes. Users go from node 1 to node 2 and out of node 2 return to node 2 exactly one time and then go to node 1, and the procedure repeats. If users are M in number:

a) find the average number of users at each node.

In the network with the users above other users arrive from outside the network at node 1 according to a Poisson flow of rate λ. These users are routed exactly as the users above, except that they leave the network at each node with probability p.

b) As in a).
c) If nodes are replaced by M/M/1 systems with the same service rate μ, draw the markovian state diagram of the occupancy of nodes in case a).
d) Verify whether a product form of the Birth-and-Death type balance the diagram in c.

P.6.22 A closed network of three cascaded nodes presents M_A users of type A and M_B users of type B. The two types of users are routed in opposite directions. Type A users have a sojourn time equal to 1 at all nodes, and Type B users have a sojourn time equal to 2 at all nodes.

a) Determine the network flow rate for each type of users;
b) the ratio M_A/M_B such that at each node we have the same average number of users of the two types.

If nodes are replaced by M/M/1 systems with the same service rate μ, in the case where we have $M = 4$ users of type A only,

c) draw the markovian state diagram of (N_1, N_2), numbers in the first and second queue;
d) Verify whether a product form of the Birth-and-Death type balance the diagram in c.

P.6.23 A network of four nodes $1, 2, 3, 4$, is such that node 1 has an infinite number of servers, node 2 has two servers and the others have one server each. All servers have a negative exponential service time at rate μ. A Poisson flow A with rate λ enters the network from node 0 and is routed according the matrix given below.

$$\begin{vmatrix} 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \end{vmatrix}$$

An additional Poisson flow B at rate $\lambda/2$ enters the network reaching nodes of the path $1, 2, 3, 4$ and then $2, 3, 4$, after which leaves the network. Find

a) the flow rate at each node;
b) the average network crossing delay of users of flow B;

c) the average network crossing delay of users that enter the network at node 1.

P.6.24 An open queuing network has three nodes, $1, 2, 3$, with a number of servers equal to respectively $1, 1, \infty$, all servers having the same rate μ. An input flow A at rate λ is deterministically routed according to the path $0-1-2-3-0$. Another flow B, still of rate λ has routing probabilities $l_{03} = 1$, $l_{31} = 1$, $l_{10} = p$, $l_{13} = 1 - p$. A third flow C of rate $\lambda/2$ presents two routing alternatives: $0-1-2-0$ or $0-3-2-3-0$.

a) Find the crossing delay of flow C for each of the two possible routing strategies;

b) determine the smallest one assuming that λ and μ are such that, at node 1, the load factor due to the first two flows, is 0.4, and that $p = 0.5$.

P.6.25 A Poisson flow of users enter at rate λ a markovian queue with one server of rate μ. Out of this, users enter a similar queue with one server at rate 2μ. At the exit of the latter system users leave the network with probability p, otherwise they reach again the first queue and the routing repeats. A second flow of rate 2λ enters the network at the second queue and is routed as the first flow. Find

a) the network crossing time and the network crossing time of the second flow;
b) the marginal and the joint distributions of the content of the two queues.

6.1 Problem Solutions - Chapter 6

P.6.1 (4.20) A user flow at rate λ enters node 1 and subsequently enters node 2. At the exit of node 2 users go back to node 1 with probability p, otherwise they leave the network. A second user flow at rate γ enters node 2 and, afterwards goes to node 1 with probability q, otherwise leaves the network. The users of this flow that exit node 1 enter node 2 with probability q, otherwise they leave the network. Assuming that in all nodes users spend an average time given by

$$V(\Gamma_i) = \frac{1}{1-\Gamma_i}$$

where Γ_i is the global flow rate at node i, find

a) the global rate Γ_i of the flow at each node i;

b) the average number of users of each type at each node.

Solution a) We have two different user classes with different flows in nodes, The first generates flows λ_i and the second flows γ_i. The overall flow rate at node i is

$$\Gamma_i = \lambda_i + \gamma_i.$$

By the flow balance at nodes we get for each class

$$\lambda_i = \frac{\lambda}{1-p}$$

$$\gamma_1 = \frac{\gamma q}{1-q^2}$$

$$\gamma_2 = \frac{\gamma}{1-q^2}$$

b) Since the time spent in nodes is the same for the two classes, the average number of users of each class is attained through the Little's results with the flow rate of the considered class.

$$E[N_1^{(\lambda)}] = \frac{\lambda_1}{1-\Gamma_1}; \qquad E[N_1^{(\gamma)}] = \frac{\gamma_1}{1-\Gamma_1};$$

$$E[N_2^{(\lambda)}] = \frac{\lambda_2}{1-\Gamma_2}; \qquad E[N_2^{(\gamma)}] = \frac{\gamma_2}{1-\Gamma_2};$$

P.6.2 (4.26) In a closed network of 5 nodes i, $i = 1, 2, 3, 4, 5$ users remain in each node an average time equal to the index of the node. The routing matrix is such that from node i they reach node 1 with probability $p = 1/2$, and with probability $1/2$ they reach the next node in sequence (or node 1 when $i = 5$). Assuming there are $M = 10$ customers in the system, find

- the flow rate at each node;
- the asymptotic probability that a user is in node i at time t.

Solution - a) We use the following formula

$$\lambda_i = \frac{\nu_i M}{\sum_{k=1}^{J} \nu_k E[V_k]}$$

where ν_i is the distribution of the routing chain. We get

$$\nu_1 = 16/31 \qquad \nu_2 = 8/31 \qquad \nu_3 = 4/31 \qquad \nu_4 = 2/31 \qquad \nu_5 = 1/31$$

$$\lambda_1 = 160/57 \qquad \lambda_2 = 80/57 \qquad \lambda_3 = 40/57 \qquad \lambda_4 = 20/57 \qquad \lambda_5 = 10/57$$

b) Here we must solve the continuous-time sojourn chain:

$$\pi_i = \frac{\nu_i E[V_i]}{\sum_k \nu_k E[V_k]}$$

We get

$$\pi_1 = 16/57 \qquad \pi_2 = 16/57 \qquad \pi_3 = 12/57 \qquad \pi_4 = 8/57 \qquad \pi_5 = 5/57$$

P.6.3 (7.4) An open network of J markovian, single-server queues with service rate μ is connected in a circle. Users enter the network at queue 1, according to a Poisson flow of rate λ_0, and leave the system immediately before queue 1 with probability p. Find

a) the maximum λ_0 allowed by network stability and the server load factor;
b) the average network delay;
c) the average number of visits to each queue.

Solution This is a Jackson's Network, and we must get the flow rate λ at each queue. The flow balance provides $\lambda p = \lambda_0$ and then $\lambda = \lambda_0/p$.

a) In order each queue to be stable we must have $\lambda/\mu < 1$, and, therefore, $\lambda_0 < \mu p$.

b) The average number of users in the net is

$$E[N] = J\frac{\rho}{1-\rho}$$

having denoted $\rho = \lambda/\mu$. From the Little's result we have

$$E[D] = J\frac{\rho}{1-\rho}\frac{1}{\lambda_0} = J\frac{1}{1-\rho}\frac{1}{\mu p}$$

c) The average number of times a user takes an entire circle is $\lambda/\lambda_0 = 1/p$. This number is a geometric variable with success probability p. Hence, the result is $1/p$.

P.6.4 (7.5) An open network of 3 markovian, single-server queues with service rates μ_i, $(i = 1, 2, 3)$ are connected in a cascade. Users enter the network at queue 1, according to a Poisson flow of rate λ. At the output of queue i they reach queue 1 with probability $1 - p_i$, while with probability p_i that enter queue $i+1$, except for queue 3 where with probability p_i they leave the network. Assuming

$$\mu_i = i \quad p_1 = 0.8, \quad p_2 = 0.5, \quad p_3 = 0.3,$$

find

a) the flow rate at each queue;

b) the maximum λ allowed by network stability;

c) The average time in each queue at each entrance and in the entire life, assuming $\lambda = 0.01$;

d) the average network delay, again with $\lambda = 0.01$.

Solution- a) The network is an open Jackson's network. Flow rates are provided by the following sysytem

$$\lambda_1 = \lambda + \lambda_1(1-p_1) + \lambda_2(1-p_2) + \lambda_3(1-p_3)$$

$$\lambda_2 = \lambda_1 p_1$$

$$\lambda_3 = \lambda_2 p_2$$

which yields

$$\lambda_1 = \frac{\lambda}{p_1 p_2 p_3}$$

$$\lambda_1 = \lambda\frac{100}{12} \qquad \lambda_2 = \lambda\frac{100}{15} \qquad \lambda_3 = \lambda\frac{100}{30}$$

b) the load factors at each queue are

$$\rho_1 = \lambda\frac{100}{12} \qquad \rho_2 = \lambda\frac{100}{30} \qquad \rho_3 = \lambda\frac{100}{90}$$

the most stringent stability constraint is at queue 1, $\rho_1 \leq 1$, which yields $\lambda_1 \leq 0.12$.

c) The sojourn time of each queue are

$$E[V_i] = \frac{1/\mu_i}{1-\rho_i}$$

$$E[V_1] = \frac{12}{11} \qquad E[V_2] = \frac{15}{29} \qquad E[V_1] = \frac{30}{89}$$

and in the entire life (generalized Little's result, and Little's result)

$$E[T_i] = \frac{E[N_i]}{\lambda} = \frac{\lambda_i}{\lambda}E[V_i].$$

The result can be derived also observing that λ_i/λ is the average number of times the user eneter node i in its life.

$$E[T_1] = \frac{100}{11}, \qquad E[T_2] = \frac{100}{29}, \qquad E[T_3] = \frac{100}{89}.$$

d)

$$E[D] = \sum_i \frac{\lambda_i}{\lambda}E[V_i] = \frac{100}{11} + \frac{100}{29} + \frac{100}{89}$$

P.6.5 (7.12) Three markovian, single-server queues with rate $\mu = 1$ are closed in a circle. Assuming there are 3 users in the network, find the rate of flow, the load factor of each queue and the content distribution of queue 1.

Solution
The network is a closed Jackson's network. From the joint solution,

$$\pi(\mathbf{n}) = B_M \psi_1(n_1)\psi_2(n_2)\dots\psi_J(n_J)$$

$$\psi_j(n_j) = \prod_{r=1}^{n_j} \frac{\nu_j}{\mu_j(r)} = \left(\frac{\nu}{\mu}\right)^{n_j}$$

being $\nu_i = \nu = 1/3$. Since $n_1 + n_2 + n_3 = M$, we see that $\pi(\mathbf{n})$ is a constant (uniform), i.e., all states have the same probability. The states (n_1, n_2, n_3) are as follows

$$(3,0,0) \quad (0,3,0) \quad (0,0,3) \quad (2,1,0) \quad (2,0,1) \quad (1,2,0) \quad (0,2,1) \quad (1,0,2)$$
$$(0,1,2) \quad (1,1,1)$$

The marginal distribution of queue 1 is

$$[4/10, 3/10, 2/10, 1/10].$$

The load factor is $1 - \pi_0 = 6/10$ and the flow rate is equal to the load factor. Alternatively, the flow rate is

$$\pi_1^{(1)}\mu_1 + \pi_2^{(1)}\mu_1 + \pi_3^{(1)}\mu_1 = 6/10.$$

P.6.6 (7.15) An open network of 3 markovian, single-server queues with service rate μ has the following routing matrix

$$\begin{vmatrix} 0 & 0.5 & 0 & 0.5 \\ 0 & 0.5 & 0.5 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0.25 & 0.25 & 0.5 & 0 \end{vmatrix}$$

where 0 is the source node and λ the source rate of the Poisson input flow. Find

a) The bottleneck queue, i.e., the one that first becomes unstable when λ increases;
b) the average network delay;
c) the average network delay of a user that enters the network at node 1;
d) the average number of time a user enters queue 1, and the same figure for users that enter the network at node 1.

Solution The network is an open Jackson's network. a) Rates are given by the following balance equations

$$\begin{cases} \lambda_1 = \lambda/2 + \lambda_1/2 + \lambda_2/2 + \lambda_3/4 \\ \\ \lambda_2 = \lambda_1/2 + \lambda_3/2 \\ \\ \lambda_3 = \lambda/2 \end{cases}$$

that yield

$$\begin{cases} \lambda_1 = 3\lambda \\ \lambda_2 = (7/4)\lambda \\ \lambda_3 = \lambda/2 \end{cases}$$

We also have

$$\rho_1 = \frac{3\lambda}{\mu}; \qquad \rho_2 = \frac{7\lambda}{4\mu}; \qquad \rho_3 = \frac{\lambda}{2\mu}$$

The bottleneck is queue 1 and we must have $\lambda < \mu/3$.

b)

$$E[D] = \sum \frac{\lambda_i}{\lambda} E[V_i] = \frac{3}{\mu - 3\lambda} + \frac{7}{4\mu - 7\lambda} + \frac{1}{2\mu - \lambda}$$

c) we have

$$E[D'] = \sum \frac{\lambda_i'}{\lambda_0'} E[V_i]$$

where the apex denotes the flows of users entering the network at node 1. Note that times $E[V_i]$ remain the same, as all users suffer the delay caused by any others. We must, then, evaluate the flow rates of users entering at node 1:

$$\begin{cases} \lambda_1' = \lambda/2 + \lambda_1'/2 + \lambda_2'/2 + \lambda_3'/4 \\ \lambda_2' = \lambda_1'/2 + \lambda_3'/2 \\ \lambda_3' = 0 \end{cases}$$

yielding

$$\begin{cases} \lambda_1' = 2\lambda \\ \lambda_2' = \lambda \\ \lambda_3' = 0 \end{cases}$$

Finally

$$E[D'] = \frac{4}{\mu - 3\lambda} + \frac{8}{4\mu - 7\lambda}$$

d) the average number of time a user enters queue i is (see the class notes)

$$e_i = \frac{\nu_i}{\nu_0} = \frac{\lambda_i}{\lambda_0} = \frac{\lambda_i}{\lambda}$$

We get in the first case $e_1 = 3$ and in the second case $e_1' = 4$.

P.6.7 (7.16) A closed network of 3 markovian single-server cascaded queues with service rates $1, 2,$ and 3 respectively has 2 users. Find the load factor and the occupancy distribution of queue 1.

Solution The network is a closed Jackson's network. The joint distribution is

$$\pi(\mathbf{n}) = B_M \psi_1(n_1)\psi_2(n_2)\ldots\psi_J(n_J)$$

$$\psi_j(n_j) = \prod_{r=1}^{n_j} \frac{\nu_j}{\mu_j(r)}$$

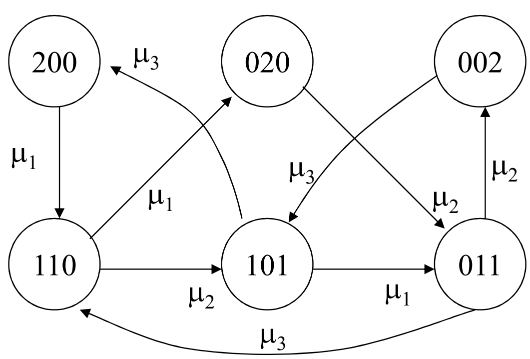
where $n_1 + n_2 + n_3 = 2,\ \mu_1(r) = 1,\ \mu_2(r) = 2,\ \mu_3(r) = 3$ and $\nu_i = 1/3$. We have

$$\pi(1,1,0) = B_2 \frac{(1/3)^2}{2} \qquad \pi(0,1,1) = B_2 \frac{(1/3)^2}{6} \qquad \pi(1,0,1) = B_2 \frac{(1/3)^2}{3}$$

$$\pi(2,0,0) = B_2 \frac{(1/3)^2}{1} \qquad \pi(0,2,0) = B_2 \frac{(1/3)^2}{2^2} \qquad \pi(0,0,2) = B_2 \frac{(1/3)^2}{3^2}$$

and normalizing, $B_2 = 324/85$.

It is also possible to derive the joint distribution by the state diagram, that in this case is



very simple,

The required distribution is

$$\alpha_0 = \pi(0,1,1) + \pi(0,0,2) + \pi(0,2,0)$$

$$\alpha_1 = \pi(1,1,0) + \pi(1,0,1)$$

$$\alpha_2 = \pi(2,0,0)$$

while, the loading factor is

$$\rho_1 = \alpha_1 + \alpha_2$$

P.6.8 (7.24) An open network of two queues with a single server at rate μ has two types of users, A and B. The first type enters the network according to a Poisson flow of rate λ_{0A}, reaches queue 1, then queue 2, then goes back to queue 1 with probability p, repeating the process, otherwise leaves the system. The second flow of B users enters queue 2 according to a Poisson flow of rate λ_{0B}, then enters queue 1 with probability q, otherwise leave the system. These users, after leaving queue 1 enter queue 2 with probability q, repeating the process, otherwise leave the system. Find

a) the flow rate at each queue and the condition on $\lambda_{0A}, \lambda_{0B}$ allowed by stability;

b) the average number of users of each type at each queue;

Solution This is not a Jackson's network because routing is not markovian. In fact the two different user types have different routing, which introduces memory. This is a Kelly's network, whose solution is shown to be the same as Kelly's.

a) We must sum the flows at each queue due to the two entering types. We get

$$\lambda_{1A} = \lambda_{2A} = \frac{\lambda_{0A}}{1-p}$$

$$\lambda_{1B} = \frac{\lambda_{0B}q}{1-q^2}$$

$$\lambda_{2B} = \frac{\lambda_{0B}}{1-q^2}$$

Since we always have $\lambda_2 \geq \lambda_1$, the stability condition is $\lambda_2 < \mu$:

$$\frac{\lambda_{0A}}{1-p} + \frac{\lambda_{0B}}{1-q^2} < \mu$$

b) The average sojourn time in the queues does not depend on the users' type and can be evaluated as usual. The average number of users of each type can be obtained by the Little's results:

$$E[N_{1A}] = \frac{\lambda_{1A}/\mu}{1-\lambda_1/\mu}; \qquad E[N_{1B}] = \frac{\lambda_{1B}/\mu}{1-\lambda_1/\mu};$$

$$E[N_{2A}] = \frac{\lambda_{2A}/\mu}{1-\lambda_2/\mu} \qquad E[N_{2B}] = \frac{\lambda_{2B}/\mu}{1-\lambda_2/\mu}$$

P.6.9 (7.22) In an open network of three queues,1,2 and 3, with single server at rates $\mu, 2\mu, \mu$, respectively a Poisson flow of rate λ enters node 1 and goes, with equal probability, toward nodes 2 and 3, and then leaves the network. Another Poisson flow of rate 2λ enters node 2 and goes, with equal probability, toward nodes 1 and 3, and then leaves the network. A third Poisson flow at rate λ enters at node 3. We have two ways to route it: A toward node 1 and then out of the network; B toward node 2 and then out of the network. Find

a) the global flow rate at nodes in the two cases A and B;

b) which routing strategy, between A and B, minimizes the network delay of the third flow.

c) which routing strategy, between A and B, minimizes the network delay.

Evaluation can be done with $\lambda = 1$ and $\mu = 5$.

Solution This is not a Jackson's network because routing is not markovian. In fact the two different user types have different routing, which introduces memory. This is a Kelly's network, whose solution is shown to be the same as Kelly's.

a) case A

$$\begin{cases} \lambda_1 = \lambda_{1A} + \lambda_{1B} = 3\lambda \\ \lambda_2 = \lambda_{2A} + \lambda_{2B} = 2.5\lambda \\ \lambda_3 = \lambda_{3A} + \lambda_{3B} = 2.5\lambda \end{cases}$$

case B

$$\begin{cases} \lambda_1 = \lambda_{1A} + \lambda_{1B} = 2\lambda \\ \lambda_2 = \lambda_{2A} + \lambda_{2B} = 3.5\lambda \\ \lambda_3 = \lambda_{3A} + \lambda_{3B} = 2.5\lambda \end{cases}$$

b) case A

$$\begin{cases} E[V_{1A}] = \dfrac{1}{\mu - 3\lambda} \\ E[V_{2A}] = \dfrac{1}{2\mu - 2.5\lambda} \\ E[V_{3A}] = \dfrac{1}{\mu - 2.5\lambda} \end{cases}$$

Case B

$$\begin{cases} E[V_{1B}] = \dfrac{1}{\mu - 2\lambda} \\ E[V_{2B}] = \dfrac{1}{2\mu - 3.5\lambda} \\ E[V_{3B}] = \dfrac{1}{\mu - 2.5\lambda} \end{cases}$$

Delays sum

$$E[D] = E[V_i] + E[V_3]$$

and we must compare $E[V_{1A}]$ with $E[V_{2B}]$. B is the best

c) we must compare

$$E[D_A] = \sum_i \frac{\lambda_{iA}}{4\lambda} E[V_i]$$

and

$$E[D_B] = \sum_i \frac{\lambda_{iB}}{4\lambda} E[V_i]$$

P.6.10 (7.23) A closed network is composed of three cascaded queues, each with a single markovian server at rate μ. Denoted by M the number of users in the network, find

a) The occupancy distribution π_i, $i = 0, 1, 2, 3$ of the first queue in the case $M = 3$;
b) The flow rate at the queues and the average time it takes for a customer to make a complete circle;
c) The average network delay of an open network with the same cascaded queues with a flow whose rate is equal to the one found in b).

Solution Closed Jackson's network.
a) The general solution in this case provides

$$\pi(\mathbf{n}) = \frac{B_M}{\mu^M}$$

showing that any state $\mathbf{n}$ has the same probability, i.e., equal to

$$\pi(\mathbf{n}) = \binom{M+J-1}{J-1}^{-1} = \binom{5}{2}^{-1} = 1/10.$$

Therefore we get

$$\pi_0 = \pi(0,0,3) + \pi(0,1,2) + \pi(0,2,1) + \pi(0,3,0) = 4/10$$
$$\pi_1 = \pi(1,0,2) + \pi(1,1,1) + \pi(1,2,0) = 3/10$$
$$\pi_2 = \pi(2,0,1) + \pi(2,1,0) = 2/10$$
$$\pi_3 = \pi(3,0,0) = 1/10$$

b) In the notes we have shown that

$$\lambda_i = \frac{\nu_i B_M}{B_{M-1}}$$

However, in our distribution we have incorporated distribution $\nu_i = 1/M$ (uniform) into B_M, which is now

$$B_M = \pi(\mathbf{n})\mu^M = \binom{M+J-1}{J-1}^{-1}\mu^M$$

Then,

$$\lambda_i = \frac{B_M}{B_{M-1}} = \binom{M+J-2}{J-1} / \binom{M+J-1}{J-1}\mu = \frac{M}{M+J-1}\mu = \frac{M}{M+2}\mu$$

Note that $\lim_{M\to\infty} = \mu$, since alle queues are always full.

The average time for a complete circle is attained with the Little's result applied to section of the network:

$$E[D] = \frac{M}{\lambda_i} = (M+2)(1/\mu)$$

(with only one user we get 3μ).

c)

$$E[D] = \frac{3}{1-\rho}(1/\mu) = \frac{3}{1 - M/(M+2)}(1/\mu) = \frac{3}{2}(M+2)(1/\mu)$$

With the open network the network delay is 1.5 times more than in the closed network.

P.6.11 In an open network of three single server queues of rate μ, users arrive at node 1 according to a Poisson flow of rate λ. Then, they are routed along the following path: $1 \to 2 \to 1 \to 3 \to 1 \to 3 \to 0$. Find the network delay.

Solution. This is a deterministic routing, with memory, and the network is not a Jackson's network. This is a Kelly's network, whose solution is shown to be the same as Kelly's.

By the routing path we have

$$\lambda_1 = 3\lambda, \qquad \lambda_2 = \lambda, \qquad \lambda_3 = 2\lambda,$$

The delay is

$$E[D] = \sum \frac{\lambda_i}{\lambda} E[V_i] = \sum \frac{\lambda_i}{\lambda} \frac{1}{\mu - \lambda_i}$$

P.6.12 (7.3) In an open network of two cascaded queues with a single server of rate μ, users arrive at queue 1 according to a Poisson process of rate λ. Users that leave queue 1 enter queue 2, and those leaving queue 2 enters queue 1 with probability p, otherwise leave the network. Draw the state diagram and show that the joint distribution is the product of the form of the distributions of the queues as if they were isolated.

Solution. The state diagram shown in Figure 6.1, part a) and b) respectively. The flow rate at the two queues is $\lambda_T = \lambda/(1-p)$. We can plug the product-form solution into balance equation and see that they are verified, as we expect, since the network is a Jackson's network. An alternative is to show that the diagram b) is equivalent to the diagram c).

The balance of c) requires

$$\pi_{00} \frac{\lambda}{1-p} = \pi_{10}\mu_1 = \pi_{01}\mu_2$$

Therefore, if diagrams b) e c) have the same solution they can be shifted as shown in d) (ithe balance is not changed). We also have

$$\pi_{00} \frac{\lambda p}{1-p} = \pi_{01}\mu_2 p$$

and also the last diagonal flux in d) can be shifted like in c), where in the transition $00 \to 10$ we sum the flux above to flux $\pi_{00}\lambda$.

P.6.13 (7.30) In an open network of two cascaded queues users arrive at queue 1 according to a Poisson process of rate λ. Users that leave queue 1 enter queue 2, and those leaving queue 2 enters queue 1 with probability $1-p$, otherwise leave the network. Queue 1 has a markovian single server of rate 2μ, while queue 2 has two markovian servers at rate μ and no waiting room and users that can not enter the queue leave the network.

a) Draw the state diagram (n_1, n_2) and check wheter the solution is the product form of the distributions of the queues in isolation.

b) Find the average network delay .

Solution (to be completed with the diagram)
The network is not a Jackson's network due to blocking and re-routing. We adopt the

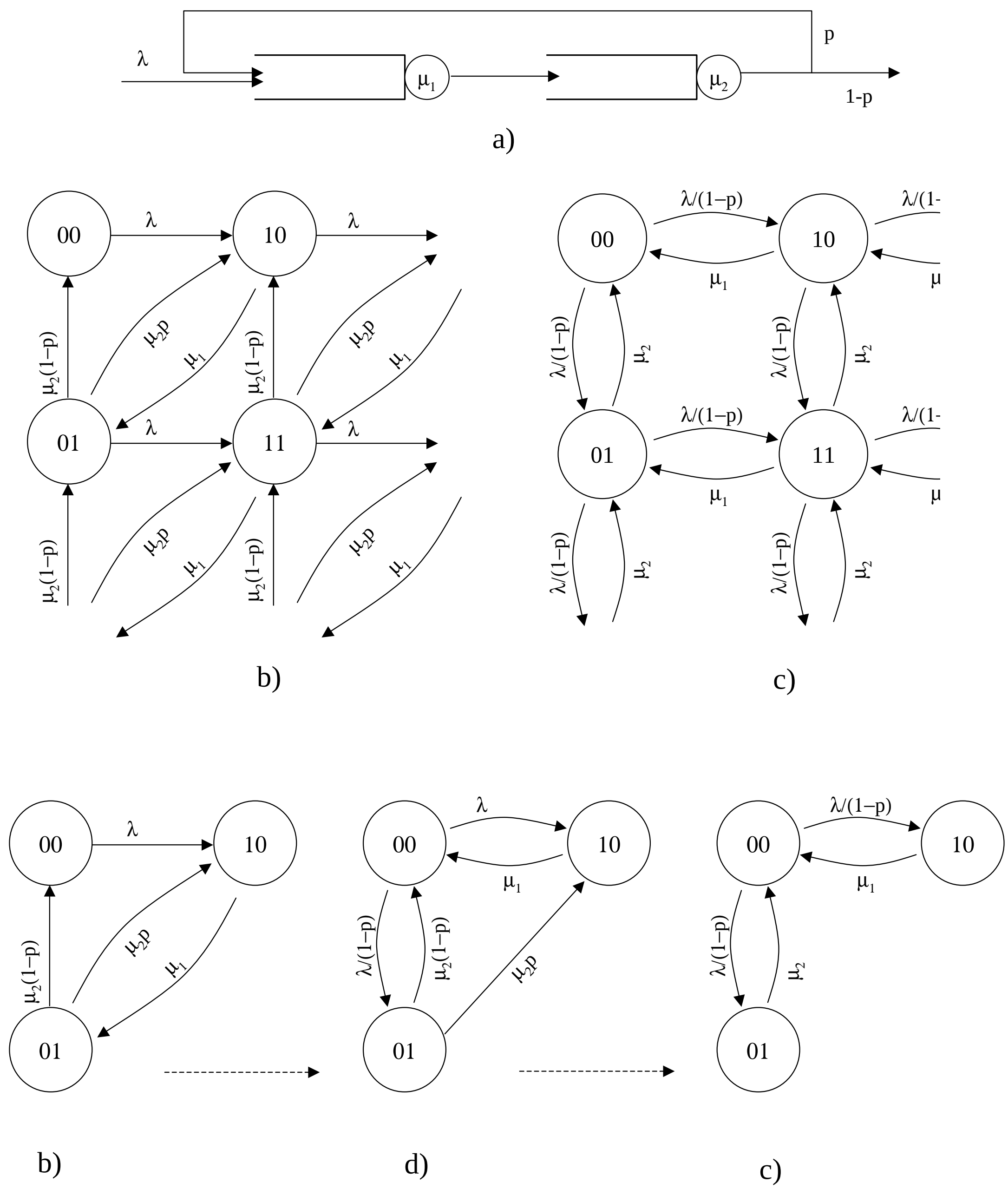


Figure 6.1:

following shortcut. If the second queue had infinite waiting room, the network would be a Jackson's network, and the distribution would be ther product form. Changing to the 0 waiting room we drop the following fluxes

$$\begin{array}{rcll} (i,2) & \longrightarrow & (i-1,3) & \text{with rate } 2\mu, \quad i \geq 1 \\ (i-1,3) & \longrightarrow & (i-1,2) & \text{with rate } 2\mu p \quad i \geq 1 \\ (i-1,3) & \longrightarrow & (i,2) & \text{with rate } 2\mu(1-p) \quad i \geq 1 \end{array}$$

while the following ones become

$$\begin{array}{rcll} (i,2) & \longrightarrow & (i,2) & \text{with rate } 2\mu(1-p), \quad i \geq 1 \\ (i,2) & \longrightarrow & (i-1,2) & \text{with rate } 2\mu p \quad i \geq 1 \end{array}$$

Hence the dropping does not alter the balance of fluxes at $(i,2)$ and the product form is still valid.

On the other side, even when the second queue is insulated the solution does not changes dropping the waiting room. So, the solution for the proposed system is of the distributions of the queues in isolation.

b)

$$E[D] = \frac{\lambda_1}{\lambda} E[V_1] + \frac{\lambda_2}{\lambda} E[V_2]$$

where

$$\lambda_1 = \lambda_2 = \lambda/p.$$

Furthermore

$$E[V_1] = \frac{1/(2\mu)}{1 - \lambda/(2p\mu)},$$

$$E[V_2] = B_2 \frac{1}{\mu}$$

where B_2 is the Erlang-B formula with two servers.

P.6.14 - A closed network of 4 markovian queues, 1,2,3,4, has $M = 100$ users and the following routing matrix

$$\begin{vmatrix} 0 & 0.5 & 0 & 0.5 \\ 0 & 0.5 & 0.5 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0 & 0 & 0.5 \end{vmatrix}$$

All the queues have an *infinite* number of servers whose rates are equal to the index of the node. Find

a) the flow rates at each queue;
b) the average delay spent in the subnetwork composed of nodes $2, 3, 4$, by a user between the exit and the return to node 1;

(Note that with an infinite number of servers there is no queue and the time in the node is equal to...)

Solution The network is a closed Jackson's network. However the solution of the problem does not need the product-form distribution since the average time spent in nodes is known, being its service time of average $1/\mu_i = 1/i$.

a) The average time $E[V_i]$ spent at each queue equals the reverse of the service rate (no queue), and, therefore, is equal to $1, 1/2, 1/3, 1/4$, respectively. The flow rates can be evaluated by

$$\lambda_i = \frac{\nu_i M}{\sum_{k=1}^{J} \nu_k E[V_k]}$$

where ν_i is the distribution of the routing chain. In our case we get

$$\nu_1 = 1/5, \qquad \nu_2 = 2/5, \qquad \nu_3 = 1/5, \qquad \nu_4 = 1/5,$$

$$\lambda_1 = 1200/31, \qquad \lambda_2 = 2400/31, \qquad \lambda_3 = 1200/31, \qquad \lambda_4 = 1200/31,$$

b) It is exactly provided as in open networks, where instead of node 0 here we have node 1.

$$D = \sum_{i \neq 1} \frac{\lambda_i}{\lambda_1} E[V_i] = 19/12$$

P.6.15 An open network of 4 markovian queues, 1,2,3,4, serves two Poisson flows of users, the first at rate λ, enters the network at node 1, while the second at rate 2λ, enters the network at node 2. The two flows have different routing matrices among nodes 0,1,2,3,4, as given below

$$\begin{vmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \end{vmatrix} \qquad \begin{vmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0.5 & 0 & 0 & 0.5 \\ 1 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0.5 & 0 \end{vmatrix}$$

Queue 2 has an *infinite* number of markovian servers, each of rate μ, while queues 1,3 and 4, have a single server at rates 2μ. Find

a) the bottleneck queue, ie, the one with highest loading factor ρ;
b) the average network crossing delay;
c) the average network crossing delay of users of the first flow;

Solution This is not a Jackson's network because routing is not markovian. In fact the two different user types have different routing, which introduces memory. This is a Kelly's network, whose solution is shown to be the same as Kelly's.
a) We have two routing matrices. One way is to solve for the flows of both separately, and sum the flows. We have

$$\lambda_1^{(1)} = (12/7)\lambda, \qquad \lambda_2^{(1)} = (8/7)\lambda, \qquad \lambda_3^{(1)} = (4/7)\lambda, \qquad \lambda_4^{(1)} = (6/7)\lambda,$$

$$\lambda_1^{(2)} = \lambda, \qquad \lambda_2^{(2)} = 2\lambda, \qquad \lambda_3^{(2)} = (3/2)\lambda, \qquad \lambda_4^{(2)} = \lambda.$$

$$\rho_1 = \frac{19}{14}\frac{\lambda}{\mu}, \qquad \rho_3 = \frac{29}{28}\frac{\lambda}{\mu}, \qquad \rho_4 = \frac{13}{14}\frac{\lambda}{\mu}.$$

The bottleneck queue is queue 1 since it presents the highest loading factor. Node 2 cannot be a bottleneck since it has an infinite number of servers.

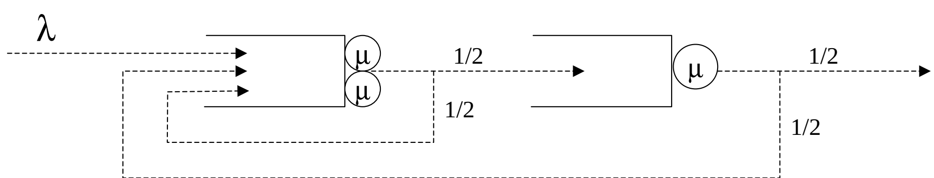
b)

$$D = \frac{19}{21}\frac{1}{2\mu(1-\rho_1)} + \frac{22}{21}\frac{1}{\mu} + \frac{29}{42}\frac{1}{2\mu(1-\rho_3)} + \frac{13}{21}\frac{1}{2\mu(1-\rho_4)}$$

c)

$$D = \frac{12}{7}\frac{1}{2\mu(1-\rho_1)} + \frac{8}{7}\frac{1}{\mu} + \frac{4}{7}\frac{1}{2\mu(1-\rho_3)} + \frac{6}{7}\frac{1}{2\mu(1-\rho_4)}$$

P.6.16 - The open network in the figure below is composed by two systems, one with two servers and the other with one server, all with service rate μ. The input flow is Poisson, with the routing probabilities shown in the figure. Find



a) the global (between birth and death) average time spent by a user in the system with one server.
b) the average number of users in the network.
c) the same as in a) and b) assuming that the input flow is dropped, so that the network becomes closed, and that there are three users in the network.

Solution (to be completed) It is a Jackson's network.
a) We easily find $\lambda_1 = 4\lambda$ and $\lambda_2 = 2\lambda$. The average number in system 2 is $\rho_2/(1-\rho_2)$ with $\rho_2 = 2\lambda/(\mu)$, and the time requested is attained dividing by $\lambda_0 = \lambda$. Therefore

$$E[D_2] = \frac{2}{\mu(1-\rho_2)}, \qquad \rho_2 = 2\lambda/\mu.$$

b) The average number in the network is provided by the sum of the average in the two systems. The average number in system 1 has been evaluated in xxx:

$$E[N_1] = \frac{2\rho_1}{1-\rho_1^2}, \qquad \rho_1 = 4\lambda/(2\mu).$$

$$E[N] = E[N_1] + E[N_2] = \frac{2\rho_1}{1-\rho_1^2} + \frac{\rho_2}{1-\rho_2}.$$

c)

P.6.17 - An open network of 3 markovian, single-server queues with service rate μ has the following routing matrix

$$\begin{vmatrix} 0 & 0.4 & 0 & 0.6 \\ 0 & 0.6 & 0.4 & 0 \\ 0.5 & 0.25 & 0 & 0.25 \\ 0 & 0.5 & 0.5 & 0 \end{vmatrix}$$

where 0 is the source node and 2λ the source rate of the Poisson input flow. A second Poisson flow of rate $\lambda/2$ enters the network at node 2, then goes round to nodes 3, 1 and back to 2. This for exactly two rounds and finally it leaves the network. Find

a) The bottleneck queue, i.e., the one that first becomes unstable when λ increases, and the corresponding maximum λ;
b) the average network crossing delay, and the average network crossing delay suffered by the second flow.
c) the average of the total time spent by a user of the second flow in node 3.

Solution This is not a Jackson's network because routing is not markovian. In fact the two different user types have different routing, which introduces memory. This is a Kelly's network, whose solution is shown to be the same as Kelly's.
a)We must distinguish the rates of the two different flows. For the first and the second we have respectively

$$\begin{cases} \lambda_1^{(1)} = \dfrac{29}{4}\lambda \\ \lambda_2^{(1)} = 4\lambda \\ \lambda_3^{(1)} = \dfrac{11}{5}\lambda \end{cases}$$

$$\begin{cases} \lambda_1^{(2)} = \lambda \\ \lambda_2^{(2)} = \dfrac{3}{2}\lambda \\ \lambda_3^{(2)} = \lambda \end{cases}$$

and the global rates

$$\begin{cases} \lambda_1 = \dfrac{33}{4}\lambda \\ \lambda_2 = \dfrac{11}{2}\lambda \\ \lambda_3 = \dfrac{16}{5}\lambda \end{cases}$$

Since the servers have the same service rate, the maximum load factor is the one with the higher flow rate, i.e., queue 1. The corresponding maximum λ allowed for stability is then $\lambda < (4/33)\mu$.

The average crossing delay is

$$E[D] = \sum \frac{\lambda_i}{\lambda_{in}} E[V_i] = \sum \frac{\lambda_i}{(5/2)\lambda} \; \frac{1}{\mu - \lambda_i}$$

and the average network crossing delay suffered by the second flow

$$E[D_2] = \sum \frac{\lambda_i^{(2)}}{\lambda_{in}^{(2)}} E[V_i] = \sum \frac{\lambda_i^{(2)}}{(1/2)\lambda} \; \frac{1}{\mu - \lambda_i}$$

By the generalized Little's result we have

$$E[T_2] = \frac{\lambda_3^{(2)}}{\lambda_{in}^{(2)}} E[V_3] = \frac{\lambda_3^{(2)}}{(1/2)\lambda} \; \frac{1}{\mu - \lambda_3} = \frac{10}{5\mu - 16\lambda_3}$$

P.6.18 - Compare the average crossing delay of the following three networks, with markovian servers at rate μ, when arrivals are Poisson at rate λ:

a) two cascaded single-server systems at rate μ;
b) one single-server system at rate μ, when after completing service users go back to the queue with probability 1/2, and otherwise leaves;
c) double server system, each server at rate $\mu/2$.

Solution

a) It is a Jackson's Network where each system behaves as in isolation. The delay is the sum of the two delays:

$$E[D] = 2\frac{1}{\mu - \lambda}$$

b) Again a Jackson's Network where $\lambda_1 = 2\lambda$ and users enter the system, on the average, two times:

$$E[D] = 2\frac{1}{\mu - 2\lambda}$$

c) It has been shown on the notes that, denoting by ρ is the server's load factor, we have

$$\pi_i = \pi_0 2\rho^i \qquad i \geq 1$$

$$\pi_0 = \frac{1-\rho}{1+\rho}$$

$$E[N] = \sum i\pi_i = 2\frac{1-\rho}{1+\rho}\sum_{i=1}^{\infty} i\rho^i = \frac{2}{1+\rho}\frac{\rho}{1-\rho} = 2\frac{\rho}{1-\rho^2}$$

where we have used the know summation

$$(1-\rho)\sum_{i=1}^{\infty} i\rho^i = \frac{\rho}{1-\rho}.$$

In our case the server's load factor is $\rho = (\lambda/2)/(\mu/2) = \lambda/\mu$, and

$$E[V] = \frac{N}{\lambda} = \frac{2}{\mu}\frac{1}{1-\rho^2}.$$

P.6.19 - A Poisson flow with rate λ enters a network from node 0 and is routed according matrix given below.

$$\begin{vmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \end{vmatrix}$$

Assuming that each user stays in each node for a time equal to T, find

a) the flow rate at each node;
b) the average number of users in the network of nodes 1,2,3,4, and its average crossing delay (between birth and death).

Answer again to questions a) and b) in the case where the number of users M in the entire network, including node 0, is M, and that the sojourn time in node 0 is M/λ (the network including node 0 is a closed network). What if $M \to \infty$?

Solution (to be completed) The sojourn time in nodes does not depend on flows.
a) The flows in the nodes are

$$\lambda_1 = (12/7)\lambda, \qquad \lambda_2 = (8/7)\lambda, \qquad \lambda_3 = (4/7)\lambda, \qquad \lambda_4 = (6/7)\lambda,$$

b) Using the Little's results we sum the average numbers of users in the different nodes

$$E[N] = \sum_i \lambda_i T = \frac{30}{7}\lambda T,$$

and the average crossing delay as

$$E[D] = E[N]/\lambda = \frac{30}{7}T$$

P.6.20 Compare the average crossing delay of the following two queuing systems, with markovian servers, when arrivals are Poisson at rate λ:

a) two cascaded double-server systems at rate μ;
b) one single-server system at rate 2μ when users after service re-enter the system with probability 1/2.

Solution Jackson's networks.
a) It has been shown on the notes and elsewhere that, for the double-server system we have

$$E[V] = \frac{1}{\mu}\frac{1}{1-\rho^2},$$

and is doubled for the given system

so the crossing time of two system is (Jackson's network)

$$E[D] = 2\frac{1}{\mu}\frac{1}{1-\rho^2}$$

b) The rate of the flow entering the M/M/1 system is 2λ, two times on the average for each users. We have

$$E[D] = 2\frac{1}{2\mu}\frac{1}{1-\rho} = \frac{1}{\mu}\frac{1}{1-\rho}, \qquad \rho = 2\lambda/2\mu.$$

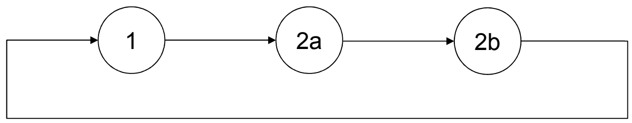
P.6.21 A closed network of two nodes is such that users have a sojourn time equal to T and $2T$ in the two nodes. Users go from node 1 to node 2 and out of node 2 return to node 2 exactly one time and then go to node 1, and the procedure repeats. If users are M in number:

a) find the average number of users at each node.

In the network with the users above other users arrive from outside the network at node 1 according to a Poisson flow of rate λ. These users are routed exactly as the users above, except that they leave the network at each node with probability p.

b) As in a).
c) If nodes are replaced by M/M/1 systems with the same service rate μ, draw the markovian state diagram of the occupancy of nodes in case a).
d) Verify whether a product form of the Birth-and-Death type balance the diagram in c.

Solution - a) the memory in returning to node 2 can be taken into account as shown below



We may use Little's result and evaluate

$$\lambda_i = \frac{\nu_i M}{\sum_{k=1}^{J} \nu_k E[V_k]}, \qquad E[N_i] = M \frac{\nu_i E[V_i]}{\sum_{k=1}^{J} \nu_k E[V_k]}.$$

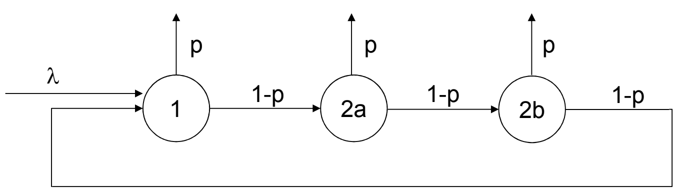
We have $\nu_i = 1/3$, and get

$$E[N_1] = M/5, \qquad E[N_{2a}] = E[N_{2b}] = 2M/5,$$

$$E[N_1] = M/5, \qquad E[N_2] = 4M/5.$$

A shortcut is observing that the time spent in node 2 is 4 times the time spent in 1, which provides 4 times users in node 2 than in node 1.

b) the added users are routed as shown below



with flow balance we get

$$\lambda_1 = \frac{\lambda}{1-(1-p)^3}, \qquad \lambda_{2a} = \frac{\lambda(1-p)}{1-(1-p)^3}, \qquad \lambda_{2b} = \frac{\lambda(1-p)^2}{1-(1-p)^3}$$

and finally

$$E[N_1] = \frac{\lambda T}{1-(1-p)^3}, \qquad E[N_2] = \frac{\lambda(1-p)2T}{1-(1-p)^3}(2-p).$$

Then these numbers must be added to those in a).

c) and d) can be answered as shown in section 6.2.5 in class notes.

P.6.22 A closed network of three cascaded nodes presents M_A users of type A and M_B users of type B. The two types of users are routed in opposite directions. Type A users have a sojourn time equal to 1 at all nodes, and Type B users have a sojourn time equal to 2 at all nodes.

a) Determine the network flow rate for each type of users;

b) the ratio M_A/M_B such that at each node we have the same average number of users of the two types.

If nodes are replaced by M/M/1 systems with the same service rate μ, in the case where we have $M = 4$ users of type A only,

c) draw the markovian state diagram of (N_1, N_2), numbers in the first and second queue;

d) Verify whether a product form of the Birth-and-Death type balance the diagram in c.

Solution a) The flow rates are given by

$$\lambda_i = \frac{\nu_i M}{\sum_{k=1}^{J} \nu_k E[V_k]},$$

and it turns out that all flows are equal since $E[V_k]$ are equal for all k. In particular we have

$$\lambda_a = M_a/3, \qquad \lambda_b = M_b/6.$$

b) The little's Result easily shows that it must be $M_a/M_b = 1$.

c) and d) are answered in section 5.2.5 of the class notes.

P.6.23 A network of four nodes $1, 2, 3, 4$, is such that node 1 has an infinite number of servers, node 2 has two servers and the others have one server each. All servers have a negative exponential service time at rate μ. A Poisson flow A with rate λ enters the network from node 0 and is routed according the matrix given below.

$$\begin{vmatrix} 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \end{vmatrix}$$

An additional Poisson flow B at rate $\lambda/2$ enters the network reaching nodes of the path $1, 2, 3, 4$ and then $2, 3, 4$, after which leaves the network. Find

a) the flow rate at each node;

b) the average network crossing delay of users of flow B;

c) the average network crossing delay of users that enter the network at node 1.

Solution

a) We have two routing mechanisms. We must solve for the flows of both separately, and sum the flows. We have

$$\lambda_1^{(A)} = (10/7)\lambda, \qquad \lambda_2^{(A)} = (9/7)\lambda, \qquad \lambda_3^{(A)} = (8/7)\lambda, \qquad \lambda_4^{(A)} = (5/7)\lambda,$$

$$\lambda_1^{(B)} = \lambda/2, \qquad \lambda_2^{(B)} = \lambda, \qquad \lambda_3^{(B)} = \lambda, \qquad \lambda_4^{(B)} = \lambda.$$

The total flow at the nodes is the summation:

$$\lambda_1 = (27/14)\lambda, \qquad \lambda_2 = (16/7)\lambda, \qquad \lambda_3 = (15/7)\lambda, \qquad \lambda_4 = (12/7)\lambda.$$

b) The average time spent at the nodes is respectively $E[V_1]$, equal to the service time since we have no queue, $E[V_2]$ is the time in the queue of 3a) plus the service time, and the remaining are those we have in M/M/1 systems. Hence the required crossing time is

$$D^{(B)} = \sum_{i=1}^{4} \frac{\lambda_i^{(B)}}{\lambda^{(B)}} E[V_i] = \frac{\lambda_1^{(B)}}{\lambda^{(B)}} \frac{1}{\mu} + \frac{\lambda_2^{(B)}}{\lambda^{(B)}} \frac{1}{\mu} \frac{\rho^2}{1-\rho^2} +$$

$$+ \frac{\lambda_3^{(B)}}{\lambda^{(B)}} \frac{1}{\mu(1-\rho_3)} + \frac{\lambda_4^{(B)}}{\lambda^{(B)}} \frac{1}{\mu(1-\rho_4)}$$

with

$$\lambda^{(B)} = \lambda/2$$

c) We must derive the flows caused by users that enter the network at node 1. This can be done by canceling other entering flows. We still have flows A and B, and proceed as above.

P.6.24 An open queuing network has three nodes, $1, 2, 3$, with a number of servers equal to respectively $1, 1, \infty$, all servers having the same rate μ. An input flow A at rate λ is deterministically routed according to the path $0-1-2-3-0$. Another flow B, still of rate λ has routing probabilities $l_{03} = 1$, $l_{31} = 1$, $l_{10} = p$, $l_{13} = 1 - p$. A third flow C of rate $\lambda/2$ presents two routing alternatives: $0-1-2-0$ or $0-3-2-3-0$.

a) Find the crossing delay of flow C for each of the two possible routing strategies;
b) determine the smallest one assuming that λ and μ are such that, at node 1, the load factor due to the first two flows, is 0.4, and that $p = 0.5$.

Solution

a) For flows A B and C we have

$$\begin{aligned}
&\lambda_1^{(A)} = \lambda, && \lambda_2^{(A)} = \lambda, && \lambda_3^{(A)} = \lambda.\\
&\lambda_1^{(B)} = \lambda/p, && \lambda_2^{(B)} = 0, && \lambda_3^{(B)} = \lambda/p.\\
&\lambda_1^{(C1)} = \lambda/2, && \lambda_2^{(C1)} = \lambda/2, && \lambda_3^{(C1)} = 0.\\
&\lambda_1^{(C2)} = 0, && \lambda_2^{(C2)} = \lambda/2, && \lambda_3^{(C2)} = \lambda.
\end{aligned}$$

With the first routing pattern we have

$$D = \frac{\lambda_1^{(C1)}}{\lambda/2} \frac{1}{\mu(1-\rho_1)} + \frac{\lambda_2^{(C1)}}{\lambda/2} \frac{1}{\mu(1-\rho_2)} + \frac{\lambda_3^{(C1)}}{\lambda/2} \frac{1}{\mu} = \frac{1}{\mu(1-\rho_1)} + \frac{1}{\mu(1-\rho_2)}$$

with

$$\rho_1 = \lambda_1/\mu = (\lambda/\mu)\frac{3p+2}{2p}, \qquad \rho_2 = \lambda_2/\mu = (\lambda/\mu)\frac{3}{2}.$$

With the second routing pattern we have

$$D' = \frac{\lambda_1^{(C2)}}{\lambda/2} \frac{1}{\mu(1-\rho_1')} + \frac{\lambda_2^{(C2)}}{\lambda/2} \frac{1}{\mu(1-\rho_2')} + \frac{\lambda_3^{(C2)}}{\lambda/2} \frac{1}{\mu} = \frac{1}{\mu(1-\rho_2')} + \frac{2}{\mu}$$

with

$$\rho_2' = \lambda_2'/\mu = (\lambda/\mu)\frac{3}{2}.$$

b) The load factor of the first two flows at node 1 is

$$\frac{\lambda_1^{(A)} + \lambda_1^{(B)}}{\mu} = 0.4, \qquad (\lambda/\mu) = \frac{2}{15}.$$

Substituting we get

$$D = \frac{1}{\mu(1-7/15)} + \frac{1}{\mu(1-3/15)} = \frac{1}{\mu}\frac{25}{8}$$

$$D' = \frac{1}{\mu(1-3/15)} + \frac{2}{\mu} = \frac{1}{\mu}\frac{26}{8}$$

We have $D < D'$.

P.6.25 A Poisson flow of users enter at rate λ a markovian queue with one server of rate μ. Out of this, users enter a similar queue with one server at rate 2μ. At the exit of the latter system users leave the network with probability p, otherwise they reach again the first queue and the routing repeats. A second flow of rate 2λ enters the network at the second queue and is routed as the first flow. Find

a) the network crossing time and the network crossing time of the second flow;
b) the marginal and the joint distributions of the content of the two queues.

Solution
The flow rates of the first flow are

$$\lambda_1^{(A)} = \lambda_2^{(A)} = \lambda/p$$

The flow rates of the second flow are

$$\begin{cases} \lambda_1^{(B)} = 2\lambda\dfrac{1-p}{p} \\ \\ \lambda_2^{(B)} = 2\lambda/p \end{cases}$$

The load factors at the two servers are respectively

$$\rho_1 = \frac{3-2p}{p}\frac{\lambda}{\mu}, \qquad \rho_2 = \frac{3}{2p}\frac{\lambda}{\mu}.$$

a) The average network delays are (CHECK)

$$E[T] = \frac{1}{\mu}\frac{3-2p}{3p}\frac{1}{1-\rho_1} + \frac{1}{\mu}\frac{1}{p}\frac{1}{1-\rho_2}$$

$$E[T'] = \frac{1}{\mu}\frac{1-p}{p}\frac{1}{1-\rho_1} + \frac{1}{\mu}\frac{1}{p}\frac{1}{1-\rho_2}$$

b) the marginal distribution of the queues are as follows

$$\pi^{(A)}(n_1) = \pi^{(A)}(0)\left(\frac{3-2p}{p}\frac{\lambda}{\mu}\right)^{n_1},$$

$$\pi^{(B)}(n_2) = \pi^{(B)}(0)\left(\frac{3}{2p}\frac{\lambda}{\mu}\right)^{n_2},$$

and the joint is

$$\pi(n_1, n_2) = \pi^{(A)}(n_1)\pi^{(B)}(n_2)$$