Queueing Systems

Lakatos, László

Queueing Systems

írta Lakatos, László

Publication date 2013

Tartalom

Queueing Systems

1. 1 Markovian queueing systems

The origin of queueing theory is connected with the collaborator of Copenhagen Telephone Company A.K. Erlang (1878-1929) who dealt with the problems of operation of telephone centers. He published his results in 1909-1922, these results mean the beginning of queueing theory.

The intensive use of telephones started at this time, the connections among subscribers were realized via telephone centers operated by hand applying operators. For the reasonable functioning it was necessary to take into account two moments. To have a great number of subscribers requiring the realization of connections as fast as possible assumes a large number of operators. But the employment of a great number of operator becomes costly and their working time is not fully used. Of course, in such case one looks for a compromise: to have such number of operators when the system is normally functioning, but minimizing the number of operators till there are acceptable response and connection times. The answer to this question is even more complicated because of the randomness of calls and the durations of conversations are also unpredictable. Erlang examined models with Poisson arrivals (the intervals between two successive calls are exponentially distributed) and the lengths of conversations are also exponentially distributed random variables. Later, it was explored that these models might be used in other areas, too. In the 30s Khinchin studied similar problems, after widening the field of application in the 50s one could see a very intensive development in these topics. One can say that if customers enter a system, there get some service and then leave, we have a queueing (or service) problem. Because of the military applications these problems appeared mainly in the Soviet Union and USA, in the English language literature the term "queueing theory" and in the Russian language literature the term "theory of mass service" was used.

In the subject we will deal with queueing problems which at least at a certain level can be investigated by analytical methods, too, and lead to solutions in closed form. At the same time, because of the size and complexity of problems appearing in the practice, usually there is no hope for their analytical description. In such case the simulation is used, after refinements these models imitate exactly enough the behaviour of real systems and give possibility to find the optimal parameters of functioning. The principal moment of simulation is the validation (verification) of the model, i.e. it is necessary to check the correctness of results obtained by means of simulation. One possibility for such purposes is the use of a simplified but exactly computable model.

The different types of queueing systems are usually described by using the notation introduced by Kendall

$$A/B/m/n.$$

The meaning of different letters in it is as follows:

- A - the type of input process of customers, it is identified by the distribution of intervals between two successive arrivals. Its usual values are: M - exponential distribution, D - constant value, E_r - Erlang distribution of order r , G - general distribution.
- B - the distribution of service time for a customer, its possible values are the same as in the case of arrivals.
- m - the number of servers in the system.
- n - there are possible two interpretations. It means either the number of waiting places, or it gives the maximum number of simultaneous customers in the system (on service and waiting for it). If there is no restriction for the waiting room, it does not appear.

These objects characterize quite well the different types of queueing systems, but they do not reflect one very important condition. This is the service discipline determining how the customers follow one another in the service process. It may be very simple (service in the order of arrivals, in inverse order, random) or more complicated (depending on the number of present customers, remaining or elapsed service time, priorities, etc.). We will see that the investigation of systems with simple probabilistic characteristics may become rather difficult because of the order and conditions of service.

1.1. 1.1 The Poisson distribution

The Poisson distribution plays a special role in the queueing systems. In the probability theory it is introduced as an approximation for the binomial distribution, but it can be derived on another way, too.

Assume that the customers enter the system at moments $t_1,\ t_2,\ t_3, \dots$ Sometimes, it is better to consider the time intervals between successive arrival moments. We require the fulfilment of the following properties:

Stationarity: the probability of k arrivals on the interval $[t, t+h)$ depends on k and h, but does not depend on t;

Markov property: the probability of event that for $[t, t+h)$ k customers enter the system does not depend on any assumption how and how many customers entered till t;

Rareness: the probability of appearance of at least two customers for a short time interval h is

$$P_{>1}(h) = o(h).$$

Furthermore, we assume that

$$P_1(h) = \lambda h + o(h),$$

where $\lambda > 0$ is a positive constant.

During a period $t+h$ k customers enter the system, this is possible on $k+1$ different ways:

1. for t enter k customers, for h none;

2. for t enter $k-1$ customers, for h one;

....

$n+1$. for t no customer enters, for h enter k ones.

By the formula of total probability

$$P_k(t+h) = \sum_{j=0}^{k} P_j(t) P_{k-j}(h).$$

Let $R_k = \sum_{j=0}^{k-2} P_j(t) P_{k-j}(h)$. Since $P_j(t) \le 1$, we have

$$R_k \le \sum_{j=0}^{k-2} P_{k-j}(h) = \sum_{s=2}^{k} P_s(h).$$

If $k \to \infty$ then

$$R_k \le \sum_{s=2}^{k} P_s(h) \le \sum_{s=2}^{\infty} P_s(h) = P_{>1}(h).$$

So, because of rareness, $P_{>1}(h) = o(h)$. Consequently,

$$P_k(t+h) = P_k(t)P_0(h) + P_{k-1}(t)P_1(h) + o(h).$$

By our assumption $P_1(h) = \lambda h + o(h)$ and

$$P_0(h) = 1 - \sum_{s=1}^{\infty} P_s(h) = 1 - P_1(h) - \sum_{s=2}^{\infty} P_s(h) = 1 - \lambda h + o(h).$$

From these equalities follows

$$P_k(t+h) = P_k(t)[1 - \lambda h] + P_{k-1}(t)\lambda h + o(h),$$

or

$$\frac{P_k(t+h) - P_k(t)}{h} = -\lambda P_k(t) + \lambda P_{k-1}(t) + \frac{o(h)}{h},$$

from which

$$P_k'(t) = -\lambda P_k(t) + \lambda P_{k-1}(t), \qquad k \geq 1.$$

Furthermore,

$$P_0(t+h) = P_0(t)P_0(h),$$

$$P_0(t+h) = P_0(t)[1 - \lambda h + o(h)],$$

$$P_0'(t) = -\lambda P_0(t),$$

from which

$$P_0(t) = Ce^{-\lambda t},$$

and, using the initial condition $P_0(0) = 1$, we get

$$P_0(t) = e^{-\lambda t}.$$

Let $P_k(t) = e^{-\lambda t} v_k(t)$, then our original system of equations may be written in the form

$$v_k'(t) = \lambda v_{k-1}(t),$$

$$v_0'(t) = 0.$$

The initial conditions are

$$v_0(0) = P_0(0) = 1,$$

$$v_k(0) = P_k(0) = 0.$$

The solution of equations is

$$v_0(t) = 1, \quad v_1(t) = \lambda t, \quad v_2(t) = \frac{(\lambda t)^2}{2!}, \ldots, v_k(t) = \frac{(\lambda t)^k}{k!}, \ldots,$$

and so

$$P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

Now we derive our additional assumption from the stationarity, Markov property and rareness. Let us consider an interval with unit length and let θ be the probability of event that for this interval no customer appears, i.e.

$$\theta = P_0(1).$$

Let us divide this interval into n equal parts. By using stationarity and Markov property

$$\theta = \left[P_0\left(\frac{1}{n}\right)\right]^n \quad \text{or} \quad P_0\left(\frac{1}{n}\right) = \theta^{1/n}.$$

The probability of event that for an interval of length k/n no customer enters

$$P_0\left(\frac{k}{n}\right) = \theta^{k/n}.$$

Let t be a nonnegative number. One can always find such integer k that

$$\frac{k-1}{n} \leq t \leq \frac{k}{n}.$$

$P_0(t)$ is nonincreasing function of t, so

$$P_0\left(\frac{k-1}{n}\right) \geq P_0(t) \geq P_0\left(\frac{k}{n}\right),$$

i.e.

$$\theta^{\frac{k-1}{n}} \geq P_0(t) \geq \theta^{\frac{k}{n}}.$$

Let $n \to \infty$, then

$$\lim_{n\to\infty} \frac{k}{n} = \lim_{n\to\infty} \frac{k-1}{n} = t.$$

From here

$$P_0(t) = \theta^t.$$

Since $P_0(t)$ is a probability, $0 \leq P_0(t) \leq 1$. Let us choose θ as $e^{-\lambda}$, then

$$P_0(t) = \exp\{-\lambda t\}.$$

Remark.

$$P_0(t+h) = P_0(t)P_0(h)$$

is a Cauchy type functional equation, $P_0(t)$ is monotonic function of t. This function has one solution if $P_0(1)$ is a fixed value.

Till the moment we did not use the property of rareness. The probability of event the interarrival time between two successive arrivals is greater than t is $e^{-\lambda t}$. So, the distribution function of interarrival time is

$$1 - e^{-\lambda t}.$$

Since during t some customers enter, we have

$$P_0(t) + P_1(t) + P_{>1}(t) = 1.$$

For small t

$$1 - e^{-\lambda t} = \lambda t + o(t),$$

so

$$P_0(t) = 1 - \lambda t + o(t),$$

consequently,

$$P_1(t) = 1 - P_0(t) - P_{>1}(t) = \lambda t + o(t).$$

1.2. 1.2 Service with waiting

Let us have n servers and customers enter according to a Poisson process with parameter λ (i.e. the interarrival time between two successive customers has exponential distribution with parameter λ). If at the arrival moment at least one server is free the customers service begins at once, if all servers are occupied it joins the waiting queue. All customers are served by one server and at a given moment one server deals with the service of one customer. If a server becomes free, immediately begins the service of next customer if there is any. The service time of a customer is an exponentially distributed random variable with distribution function

$$B(x) = 1 - e^{-\mu x}.$$

The basic property of exponential distribution. The exponential distribution has the memoryless property, i.e. the remaining service time does not depend on how long the service has been continued.

Let $f_a(t)$ the probability of event that a service continuing already a time will still be continued at least t time. Then

$$f_0(t) = e^{-\mu a}, \qquad f_0(a+t) = e^{-\mu(a+t)}.$$

Since

$$f_0(a+t) = f_0(a) f_a(t),$$

we have

$$e^{-\mu(a+t)} = e^{-\mu a} f_a(t),$$

from which

$$f_a(t) = e^{-\mu t} = f_0(t).$$

The system of equations describing the functioning of the system. First we find the probability of event at moment $t + h$ all servers are free. We have different possibilities:

- at t all servers were free and for h no customer entered;
- at t one server was occupied, during h the service was completed and new customers did not enter;
- the probabilities of further events (2, 3,... customers were served and their services were completed) have probability of order $o(h)$.

We have

$$P_0(t)e^{-\lambda h} = P_0(t)[1 - \lambda h + o(h)],$$

$$P_1(t)e^{-\lambda h}[1 - e^{-\mu h}] = P_1(t)\mu h + o(h),$$

by means of them

$$P_0(t+h) = P_0(t)[1 - \lambda h] + P_1(t)\mu h + o(h)$$

or

$$P_0'(t) = -\lambda P_0(t) + \mu P_1(t).$$

Let us consider the case $1 \le k < n$. We have the following possibilities:

- At moment t there are k customers in the system, during h no new customer enters and the service of none of customers is terminated. The probability of this event is

$$P_k(t)e^{-\lambda h}(e^{-\mu h})^k = P_k(t)[1 - \lambda h - k\mu h + o(h)].$$

- At the moment t there are $k-1$ customers on service, during h a new customer enters and the service of none of customers is terminated. The probability of this event is

$$P_{k-1}(t)[1 - e^{-\lambda h}](e^{-\mu h})^{k-1} = \lambda h P_{k-1}(t) + o(h).$$

- At moment t there are $k+1$ customers on service, new customers do not enter and the service of a customer is completed. The probability of this event is

$$P_{k+1}(t)e^{-\lambda h}C_{k+1}^1(e^{-\mu h})^k(1 - e^{-\mu h}) = P_{k+1}(t)(k+1)\mu h + o(h).$$

The probability of further possibilities is $o(h)$.

Similarly to the determination of $P_0(t)$, from the above three possibilities we obtain the equation

$$P_k'(t) = \lambda P_{k-1}(t) - (\lambda + k\mu)P_k(t) + (k+1)\mu P_{k+1}(t).$$

Finally, let $k \ge n$. First, let us consider the case $k > n$. We have the following possibilities:

- At moment t there are $k-1$ customers in the system, during h one new customer appears and none of n services is terminated. The probability of this event is

$$P_{k-1}(t)(1 - e^{-\lambda h})(e^{-\mu h})^n = P_{k-1}(t)\lambda h + o(h).$$

- At moment t k customers are present, new customer does not enter and none of services is terminated. The probability of this event is

$$P_k(t)e^{-\lambda h}(e^{-\mu h})^n = P_k(t)[1 - \lambda h - n\mu h + o(h)].$$

- At t there are $k+1$ customers in the system. During h no new customer enters and the none of n services in process is terminated. The probability of this event is

$$P_{k+1}(t)e^{-\lambda h}C_n^1(e^{-\mu h})^{n-1}(1 - e^{-\mu h}) = P_{k+1}(t)n\mu h + o(h).$$

On the basis of three possibilities we have

$$P_k(t+h) = P_{k-1}(t)\lambda h + P_k(t)[1-\lambda h - n\mu h] + P_{k+1}(t)n\mu h + o(h),$$

from which

$$P_k'(t) = \lambda P_{k-1}(t) - (\lambda + n\mu)P_k(t) + n\mu P_{k+1}(t).$$

In the case $k=n$ the transition $k-1 \to k$ differs from the case $k>n$. The corresponding probability ($n-1$ customers are present, one new customer enters and none of $n-1$ services is terminated) is

$$P_{n-1}(t)(1-e^{-\lambda h})(e^{-\mu h})^{n-1} = P_{n-1}(t)\lambda h + o(h)$$

it coincides with the probability for the case $k>n$. So, in the case $k \geq n$ our differential equation has the form

$$P_k'(t) = \lambda P_{k-1}(t) - (\lambda + n\mu)P_k(t) + n\mu P_{k+1}(t).$$

The stationary distribution. Assume that the system has nontrivial distribution as $t \to \infty$. In this case the probabilities $P_k(t)$ tend to constant limiting values $P_k > 0$, P_k' are the derivatives of constants, so they are equal to 0. From the system of differential equations

$$\begin{aligned}
P_0'(t) &= -\lambda P_0(t) + \mu P_1(t), \\
P_k'(t) &= \lambda P_{k-1}(t) - (\lambda + k\mu)P_k(t) + (k+1)\mu P_{k+1}(t), \quad 1 \leq k < n, \\
P_k'(t) &= \lambda P_{k-1}(t) - (\lambda + n\mu)P_k(t) + n\mu P_{k+1}(t), \quad k \geq n
\end{aligned}$$

we obtain the system of linear equations

$$\begin{aligned}
-\lambda P_0 + \mu P_1 &= 0, \\
\lambda P_{k-1} - (\lambda + k\mu)P_k + (k+1)\mu P_{k+1} &= 0, \quad 1 \leq k < n, \\
\lambda P_{k-1} - (\lambda + n\mu)P_k + n\mu P_{k+1} &= 0, \quad k \geq n
\end{aligned}$$

and the normalizing condition is

$$\sum_{k=0}^{\infty} P_k = 1.$$

Let us denote

$$\begin{aligned}
z_k &= \lambda P_{k-1} - k\mu P_k, \qquad 1 \leq k < n, \\
z_k &= z_k = \lambda P_{k-1} - n\mu P_k, \qquad k \geq n.
\end{aligned}$$

By using this notation, our system of equations takes on the form

$$z_1 = 0, \qquad z_k - z_{k+1} = 0, \quad k \geq 1,$$

from which

$$z_k = 0.$$

Consequently, in the case $1 \leq k < n$ we get

$$k\mu P_k = \lambda P_{k-1}, \tag{1}$$

and in the case $k \geq n$ we obtain

$$n\mu P_k = \lambda P_{k-1}.$$

Let us introduce the notation $\rho = \frac{\lambda}{\mu}$, then from (1)

$$P_k = \frac{\rho^k}{k!} P_0$$

if $1 \leq k \leq n$ and

$$P_k = \left(\frac{\rho}{n}\right)^{k-n} P_n = \frac{\rho^k}{n! n^{k-n}} P_0$$

if $k \geq n$. (In the case $k = n$ the two formulas coincide.) These formulas express probabilities P_k via P_0. To find P_0 we have the condition

$$P_0 \left[\sum_{k=0}^{n} \frac{\rho^k}{k!} + \frac{n^n}{n!} \sum_{k=n+1}^{\infty} \left(\frac{\rho}{n}\right)^k \right] = 1.$$

The first summand in the brackets is finite, the second summand is convergent if $\frac{\rho}{n} < 1$.

We find the probability that the waiting time γ exceeds t. Let us denote this probability by $P(\gamma \geq t)$ and let $P_k(\gamma \geq t)$ be the probability of same event on condition that there are k customers in the system. Obviously,

$$P(\gamma \geq t) = \sum_{k=n}^{\infty} P_k \cdot P_k(\gamma \geq t),$$

where $P_k \quad (k = 0, 1, 2, \ldots)$ are the above computed stationary probabilities.

The service time of a customer is an exponentially distributed random variable with parameter μ. The existence of waiting queue means that the simultaneous service of n customers is in process. The probability that the service time of a customer is longer than t is $e^{-\mu t}$, and in case of n customers none of the services is finished has the probability

$$(e^{-\mu t})^n = e^{-n\mu t}.$$

Consequently, the probability of event that at least one service is terminated before t (i.e. the distribution function of time between two service time terminations) is

$$1 - e^{-n\mu t}.$$

In order to start the service of a new entering customer till t, it is necessary to begin the services of $k-n$ waiting customers and the that of actual one, i.e. till the moment t $k-n+1$ services should be terminated. The distribution of time till the realization of this event is the $k-n+1$-fold convolution of exponential distributions with parameter $n\mu$

$$1 - e^{-n\mu t} - \ldots - \frac{(n\mu t)^{k-n}}{(k-n)!} e^{-n\mu t}.$$

So, the waiting time is greater than t with probability

$$P_k(\gamma t) = \sum_{s=0}^{k-n} \frac{(n\mu t)^s}{s!} e^{-n\mu t}.$$

By using this probability, we obtain

$$\begin{aligned} P(\gamma \geq t) &= P_n e^{-n\mu t} \sum_{k=n}^{\infty} \left(\frac{\rho}{n}\right)^{k-n} \sum_{s=0}^{k-n} \frac{(n\mu t)^s}{s!} = \\ &= P_n e^{-n\mu t} \sum_{s=0}^{\infty} \frac{(n\mu t)^s}{s!} \sum_{n=k+s}^{\infty} \left(\frac{\rho}{n}\right)^{k-n} = \\ &= P_n e^{-n\mu t} \sum_{s=0}^{\infty} \frac{(n\lambda t)^s}{s! n^s} \sum_{k=n+s}^{\infty} \left(\frac{\rho}{n}\right)^{k-n-s} = \\ &= \frac{P_n}{1-\frac{\rho}{n}} e^{-n\mu t} \sum_{s=0}^{\infty} \frac{(\lambda t)^s}{s!} = \frac{P_n}{1-\frac{\rho}{n}} e^{-(n\mu-\lambda)t}. \end{aligned}$$

Let π be the probability of event that all servers are busy

$$\pi = \sum_{k=n}^{\infty} P_k = \sum_{k=n}^{\infty} P_n \left(\frac{\rho}{n}\right)^{k-n} = P_n \frac{1}{1-\frac{\rho}{n}},$$

from which

$$P_n = \pi \left(1 - \frac{\rho}{n}\right),$$

and so

$$P(\gamma \geq t) = \pi e^{-(n\mu-\lambda)t}.$$

Since

$$P(\gamma \leq t) = 1 - \pi e^{-(n\mu-\lambda)t},$$

the corresponding density function is

$$\pi(n\mu - \lambda) e^{-(n\mu-\lambda)t},$$

the mean value of waiting time

$$\int_0^{\infty} t\pi(n\mu - \lambda) e^{-(n\mu-\lambda)t} dt = \frac{\pi}{n\mu - \lambda}.$$

1.3. 1.3 Further Markov type systems

In this part we will describe queueing systems with Poisson arrivals and exponentially distributed service time.

1.3.1. 1.3.1 System with pure refusals

Let us consider a system with n servers. If at the moment of arrival there is free server then the new customers service begins at once, in the opposite case it will be lost. The system has $n+1$ different states:

- all servers are free;

- one server is occupied;

....

- all servers are occupied.

We derive the equations describing the system states. Let $p_k(t)$ be the probability of event that at moment t there are k customers in the system and consider the system at the moment $t+h$:

- In the case $k=0$ at the moment t the system is free and during h no new customer enters or at t there is one customer on service, it is terminated till $t+h$, the probability of all further events is $o(h)$. We have

$$p_0(t+h) = p_0(t)(1-\lambda h) + p_1(t)\mu h + o(h),$$

  from which

$$p_0'(t) = -\lambda p_0(t) + \mu p_1(t).$$

- In case $0<k<n$ at moment t either there are $k-1$ customers in the system and for h a new customer enters; or k customers are present and for h no new customer enters and no service is terminated; or $k+1$ customers are on service and for h a service is completed. The probability of further events is $o(h)$. We have

$$p_k(t+h) = p_{k-1}(t)\lambda h + p_k(t)[1-(\lambda+k\mu)h] + p_{k+1}(t)(k+1)\mu h + o(h),$$

  from which

$$p_k'(t) = \lambda p_{k-1}(t) - (\lambda+k\mu)p_k(t) + (k+1)\mu p_{k+1}(t).$$

- In the case $k=n$ at moment t there are $n-1$ customers in the system and for h a new customer enters, or n customers are present and the service of no customer is terminated. We have

$$p_n(t+h) = p_{n-1}(t)\lambda h + p_n(t)(1-n\mu h) + o(h),$$

  from which

$$p_n'(t) = \lambda p_{n-1}(t) - n\mu p_n(t).$$

Similarly to the previous systems at $t\to\infty$ we obtain the following system of algebraic equations:

$$\begin{aligned} -\lambda p_0 + \mu p_1 &= 0, \\ \lambda p_{k-1} - (\lambda+k\mu)p_k + (k+1)\mu p_{k+1} &= 0 \qquad (0<k<n), \\ \lambda p_{n-1} - n\mu p_n &= 0. \end{aligned}$$

This system of equations can be solved on the following way:

$$p_1 = \frac{\lambda}{\mu}p_0, \qquad p_2 = \frac{1}{2\mu}[-\lambda p_0 + (\lambda+\mu)p_1] = \frac{1}{2\mu}\left[-\lambda p_0 + \frac{\lambda^2}{\mu}p_0 + \lambda p_0\right] = \frac{\lambda^2}{2\mu^2}p_0, \ldots,$$

$$p_k = \frac{\lambda^k}{k!\mu^k}p_0, \ldots,$$

or, by introducing the notation $\rho = \lambda/\mu$,

$$p_k = \frac{\rho^k}{k!} p_0.$$

By using the condition

$$\sum_{k=0}^{n} p_k = p_0 \sum_{k=0}^{n} \frac{\rho^k}{k!} = 1$$

we get

$$p_0 = \left[\sum_{k=0}^{n} \frac{\rho^k}{k!} \right]^{-1}$$

and

$$p_k = \frac{\frac{\rho^k}{k!}}{\sum_{i=0}^{n} \frac{\rho^i}{i!}} \qquad k = 0, 1, \ldots, n.$$

These expressions are the so-called Erlang-Sevastanov formulas. Erlang derived them for the case of exponentially distributed service time, Sevastanov showed that they remain valid for the case of arbitrary service time distribution.

Example. A telephone center has 4 lines. If all lines are busy the new calls are refused. The intensity of calls is $\lambda = 3$ (calls/minute). The average duration of a conversation is 2 minutes. Find the probability of refusal and the probability that all lines are free.

The average length of a conversation is 2 minutes, so the intensity of service is $\mu = 0.5$ conversations/minute; $\rho = \lambda/\mu = 6$.

$$\begin{aligned} P_{ref} &= P_4 = \frac{\frac{\rho^4}{4!}}{1 + \rho + \frac{\rho^2}{2!} + \frac{\rho^3}{3!} + \frac{\rho^4}{4!}} \approx 0.47, \\ P_0 &= \frac{1}{1 + \rho + \frac{\rho^2}{2!} + \frac{\rho^3}{3!} + \frac{\rho^4}{4!}} \approx 0.0087. \end{aligned}$$

1.3.2. 1.3.2 Limited waiting queue

We consider a queueing system with finite waiting room. If at the moment of arrival all servers are occupied then the customers join a waiting queue on condition that there are less than m customers. If all waiting places are occupied the customer is lost. The number of servers is n, the arrival process is of Poisson type, the service time has exponential distribution with parameter μ.

The possible states are the following ones:

X_0 - all servers are free and there is no waiting customer;

X_1 - one server is busy;

....

X_k - k servers are busy;

....

X_{n-1} - $n-1$ servers are busy;

X_n - all servers are busy;

X_{n+1} - all servers are busy and there is a waiting customer;

....

X_{n+m} - all servers are busy and all waiting places are occupied.

The equations for the probabilities $p_0(t), \dots, p_{n-1}(t)$ coincide with the equations at the derivation of Erlang formulas. We find the equations for further states.

In case X_n

$$p_n(t+h) = p_{n-1}(t)\lambda h + p_n(t)[1 - \lambda h - n\mu h] + p_{n+1}(t)n\mu h + o(h),$$

from which

$$p'_n(t) = \lambda p_{n-1}(t) - (\lambda + n\mu)p_n(t) + n\mu p_{n+1}(t).$$

In case X_{n+s} $(1 \leq s < m)$

$$p_{n+s}(t+h) = p_{n+s-1}(t)\lambda h + p_{n+s}(t)[1 - \lambda h - n\mu h] + p_{n+s+1}(t)n\mu h + o(h),$$

from which

$$p'_{n+s}(t) = \lambda p_{n+s-1}(t) - (\lambda + n\mu)p_{n+s}(t) + n\mu p_{n+s+1}(t).$$

Finally, in the case $s = m$,

$$p_{n+m}(t+h) = p_{n+m-1}(t)\lambda h + p_{n+m}(t)(1 - n\mu h) + o(h),$$

from which

$$p'_{n+m}(t) = \lambda p_{n+m-1}(t) - n\mu p_{n+m}(t).$$

Collecting the all above equations the functioning of system is described by the system of equations

$$\begin{aligned}
p'_0(t) &= -\lambda p_{k-1}(t) + \mu p_1(t), \\
& \qquad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
p'_k(t) &= \lambda p_{k-1}(t) - (\lambda + k\mu)p_k(t) + (k+1)\mu p_{k+1}(t) \qquad (0 < k \leq n-1), \\
& \qquad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
p'_n(t) &= \lambda p_{n-1}(t) - (\lambda + n\mu)p_n(t) + n\mu p_{n+1}(t), \\
& \qquad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
p'_{n+s}(t) &= \lambda p_{n+s-1}(t) - (\lambda + n\mu)p_{n+s}(t) + p_{n+s+1}(t) \qquad (1 \leq s < m), \\
& \qquad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
p'_{n+m}(t) &= \lambda p'_{n+m-1}(t) - n\mu p_{n+m}(t).
\end{aligned}$$

As we have seen earlier the stationary probabilities as $t \to \infty$ can be found from the system of linear equations:

$$
\begin{aligned}
-\lambda p_0 + \mu p_1 &= 0, \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\
\lambda p_{k-1} - (\lambda + k\mu)p_k + (k+1)\mu p_{k+1} &= 0 \qquad (0 < k \le n-1), \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\
\lambda p_{n-1} - (\lambda + n\mu)p_n + n\mu p_{n+1} &= 0, \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\
\lambda p_{n+s-1} - (\lambda + n\mu)p_{n+s} + n\mu p_{n+s+1} &= 0 \qquad (1 \le s < m), \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\
\lambda p_{n+m-1} - n\mu p_{n+m} &= 0, \\
\sum_{k=0}^{n+m} p_k &= 1.
\end{aligned}
$$

The solution of this system is

$$
\begin{aligned}
p_k &= \frac{\frac{\rho^k}{k!}}{\sum_{k=0}^{n} \frac{\rho^k}{k!} + \frac{\rho^n}{n!} \sum_{s=1}^{m} \left(\frac{\rho}{n}\right)^s} \qquad (0 \le k \le n), \\
p_{n+s} &= \frac{\frac{\rho^n}{n!} \left(\frac{\rho}{n}\right)^s}{\sum_{k=0}^{n} \frac{\rho^k}{k!} + \frac{\rho^n}{n!} \sum_{s=1}^{m} \left(\frac{\rho}{n}\right)^s} \qquad (1 \le s \le m).
\end{aligned}
$$

Example. The cars arrive at a service station according to a Poisson process with parameter $\lambda = 0.5$ (cars/hour). In the workshop there is one working place and there are 3 waiting places in the yard. The average repair time of a car is $m_j = 1/\mu = 2$ (hours). With what probability the repair of a car is refused? With what probability the service station is idle? How do these values change if one new working place is added?

We have the following parameters: $\lambda = 0.5$, $\mu = 0.5$, $\rho = \lambda/\mu = 1$, $m = 3$. In the case $n = 1$ the probability of refusal is

$$p_{1+3} = \frac{1}{1+1+3} = 0.2.$$

The service station is idle with probability

$$p_0 = \frac{1}{5} = 0.2.$$

In the case $n = 2$

$$p_{2+3} = \frac{\frac{1}{16}}{1 + 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16}} \approx 0.021,$$

the probability of free state will be much higher, namely

$$p_0 = \frac{16}{47} \approx 0.34,$$

so the probability of refusal will significantly, ten times be decreased, but the average idle time will be higher more than one and a half times.

1.3.3. 1.3.3 Az $M/M/1$ system

The queueing system having the Kendall notation M/M/1 is the simplest one, both the interarrival and service times have exponential distributions with parameters λ and μ, respectively. There is one server and there is no restriction on the waiting room. Because of the simple structure its investigation leads to well interpreted characteristics and gives possibility to check results obtained for more complex systems.

We derive the system of equations describing its functioning. In case $k=0$

$$p_0(t+h) = p_0(t)[1-\lambda h] + p_1(t)\mu h + o(h),$$

from which

$$p_0'(t) = -\lambda p_0(t) + \mu p_1(t).$$

In case $k \geq 1$

$$p_k(t+h) = p_{k-1}(t)\lambda h + p_k(t)[1-\lambda h - \mu h] + p_{k+1}(t)\mu h + o(h),$$

from which

$$p_k'(t) = \lambda p_{k-1}(t) - (\lambda+\mu)p_k(t) + \mu p_{k+1}(t) \qquad (k \geq 1).$$

Following the usual way we obtain the system of algebraic equations

$$\begin{aligned} -\lambda p_0 + \mu p_1 &= 0, \\ \lambda p_{k-1} - (\lambda+\mu)p_k + \mu p_{k+1} &= 0 \qquad (k \geq 1), \\ \sum_{k=0}^{\infty} p_k &= 1. \end{aligned}$$

Its solution is

$$\begin{aligned} p_1 &= \frac{\lambda}{\mu} p_0, \\ \mu p_2 &= (\lambda+\mu)p_1 - \lambda p_0 = \frac{\lambda^2}{\mu} p_0 \quad \rightarrow \quad p_2 = \left(\frac{\lambda}{\mu}\right)^2 p_0, \\ p_3 &= \frac{\lambda+\mu}{\mu} p_2 - \frac{\lambda}{\mu} p_1 = \frac{\lambda+\mu}{\mu}\frac{\lambda^2}{\mu^2} p_0 - \lambda^2\mu^2 p_0 = \left(\frac{\lambda}{\mu}\right)^3 p_0, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ p_k &= \left(\frac{\lambda}{\mu}\right)^k p_0 = p_0 \rho^k, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned}$$

and using the normalizing condition, we obtain

$$1 = p_0 \sum_{k=0}^{\infty} \rho^k = \frac{p_0}{1-\rho} \quad \rightarrow \quad p_0 = 1-\rho.$$

So,

$$p_k = (1-\rho)\rho^k, \qquad k = 0, 1, 2, \ldots,$$

i.e. we come to the geometrical distribution.

We find the mean value of waiting time in the system. It will be the mean value of number of customers in the system multiplied by the mean value of service time for one customer:

$$t_{waiting} = \frac{1}{\mu}\sum_{k=1}^{\infty} k\cdot(1-\rho)\rho^k = \frac{(1-\rho)\rho}{\mu}\sum_{k=1}^{\infty} k\rho^{k-1} = \frac{(1-\rho)\rho}{\mu}\frac{1}{(1-\rho)^2} =$$

$$= \frac{\rho}{\mu}\frac{1}{1-\rho} = \frac{\lambda}{\mu^2}\frac{1}{1-\frac{\lambda}{\mu}} = \frac{\rho}{\mu-\lambda}.$$

2. 2 The M/G/1 system

The memoryless property of exponential distribution simplifies the examination of queueing systems, there is no necessity to consider what has happened before the actual moment. One can assume the functioning of system begins at this moment with given initial conditions, knowing the actual state it is not necessary to deal with the past. This approach was quite good for the telephone centers, but the further applications required more sophisticated investigation. Of course, it would be ideal to have a general solution of G/G/... type systems, but because of their complexity it is impossible. So, as next step, the exponentiality of service times or interarrival times was canceled, this fact led to the M/G/1 and G/M/1 systems. In this chapter we will deal with a possible approach to the M/G/1 system.

The complete description of M/G/1 system is given by the Takács integro-differential equation. A Poisson process with parameter λ enters the system, the service time of a customer is an arbitrary distributed random variable with distribution function $B(x)$. In the system there is one server and there is no restriction on the waiting room.

Let $\gamma(t)$ denote the virtual waiting time, i.e. such amount of time is required for the service of present customers. Let

$$F(t,x) = P\{\gamma(t) \le x\}.$$

The Takács integro-differential equation concerns the virtual waiting time, it has the form

$$\frac{\partial F(t,x)}{\partial t} = \frac{\partial F(t,x)}{\partial x} - \lambda F(t,x) + \lambda\int_0^x B(x-y)d_yF(t,y).$$

Its solution is possible by the double use of Laplace-Stieltjes transform. It is quite complicated, so we will use another method. In case of solution applying the Laplace-Stieltjes transform it is necessary to know the probability $F(t,0)$, i.e. the probability of free state as the function of time. In case of differential equations such boundary conditions usually follow from real physical restrictions, here it is an additional problem. To find it, the distribution function of busy period is required which itself is an interesting task.

2.1. 2.1 Solution with embedded Markov chain

The Takács integro-differential equation describes the state of M/G/1 system at any moment t, but from practical viewpoint its solution is problematic. If we do not need full information, it would be enough to examine the states only at certain moments. This approach led to the method of embedded Markov chains when we consider our system at moments having the Markov property. If we characterize the M/G/1 system by means of the number of present customers then such moments will be the termination points of services for the different customers.

Let t_n be the moment when the service of the n-th customer is terminated, N_{t_n} the number of customers left in the system after having completed the service. Then

$$N_{t_{n+1}} = \begin{cases} \Delta_n & \text{if } N_{t_n} = 0, \\ N_{t_n} + \Delta_n - 1 & \text{if } N_{t_n} \geq 1, \end{cases}$$

where Δ_n is the number of customers entering during $[t_n, t_{n+1})$. After having served the n -th customer there are N_{t_n} ones, this number is increased by the customers entering for the service of $n+1$ -st one and is decreased by the served $n+1$ -st one. In the case $N_{t_n} = 0$ after the service of n -th customer the system becomes free, after the free period enters the $n+1$ -st customer, the next state of the system is determined by the number of customers arriving during its service. In both cases the system state at t_{n+1} is determined by the system state at t_n (present customers) and the number of entering customers after t_n which, since we have Poisson arrival process, does not depend on the prehistory. So, the random variables N_{t_n} constitute a Markov chain.

We find the transition probabilities of the Markov chain. Let

$$a_j = \int_0^\infty \frac{(\lambda x)^j}{j!} e^{-\lambda x} dB(x)$$

be the probability of event that for the service of a customer j new customers arrive in the system. The matrix of transition probabilities has the form

$$\begin{bmatrix} a_0 & a_1 & a_2 & a_3 & \dots \\ a_0 & a_1 & a_2 & a_3 & \dots \\ 0 & a_0 & a_1 & a_2 & \dots \\ 0 & 0 & a_0 & a_1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}$$

The system of equations describing the ergodic distribution is

$$p_i = \sum_j p_j p_{ji},$$

in our case it takes on the form

$$\begin{aligned} p_0 &= p_0 a_0 + p_1 a_0, \\ p_k &= \sum_{j=1}^{k+1} p_j a_{k-j+1} + p_0 a_k, \qquad k \geq 1. \end{aligned}$$

Multiply the equation for p_k $(k \geq 1)$ by z^k and sum up by k :

$$\begin{aligned}
\sum_{k=1}^{\infty} p_k z^k &= \sum_{k=1}^{\infty} z^k \sum_{j=1}^{k+1} p_j a_{k-j+1} + p_0 \sum_{k=1}^{\infty} a_k z^k = \\
&= \sum_{k=1}^{\infty} z^k \sum_{j=1}^{k} p_j a_{k-j+1} + \sum_{k=1}^{\infty} z^k p_{k+1} a_0 + p_0 \sum_{k=1}^{\infty} a_k z^k = \\
&= \frac{1}{z} \sum_{k=1}^{\infty} \sum_{j=1}^{k} p_j z^j a_{k-j+1} z^{k-j+1} + \frac{1}{z} a_0 \sum_{k=1}^{\infty} p_{k+1} z^{k+1} + p_0 \sum_{k=1}^{\infty} a_k z^k = \\
&= \frac{1}{z} \sum_{j=1}^{\infty} p_j z^j \sum_{k=j}^{\infty} a_{k-j+1} z^{k-j+1} + \frac{1}{z} a_0 \sum_{k=1}^{\infty} p_{k+1} z^{k+1} + p_0 \sum_{k=1}^{\infty} a_k z^k.
\end{aligned}$$

Adding the equation for p_0, we have

$$\sum_{k=0}^{\infty} p_k z^k = \frac{1}{z} \sum_{j=1}^{\infty} p_j z^j \sum_{k=j}^{\infty} a_{k-j+1} z^{k-j+1} + \frac{1}{z} a_0 \sum_{k=0}^{\infty} p_{k+1} z^{k+1} + p_0 \sum_{k=0}^{\infty} a_k z^k$$

or

$$P(z) = \frac{1}{z}[P(z) - p_0][A(z) - a_0] + \frac{1}{z} a_0 [P(z) - p_0] + p_0 A(z),$$

where

$$P(z) = \sum_{k=0}^{\infty} p_k z^k \qquad \text{and} \qquad A(z) = \sum_{k=0}^{\infty} a_k z^k.$$

The right side of the above equation is

$$\frac{1}{z} P(z) A(z) - \frac{1}{z} A(z) p_0 - \frac{1}{z} a_0 P(z) + \frac{1}{z} p_0 a_0 + \frac{1}{z} a_0 P(z) - \frac{1}{z} p_0 a_0 + p_0 A(z),$$

so

$$zP(z) = P(z)A(z) - p_0 A(z) + p_0 z A(z),$$

from which

$$P(z) = \frac{(z-1)A(z)}{z - A(z)} p_0.$$

This expression contains the unknown probability p_0, one can find it from the condition $P(1) = 1$:

$$p_0 = \lim_{z \to 1} \frac{z - A(z)}{(z-1)A(z)} = \lim_{z \to 1} \frac{1 - A'(z)}{A(z) + zA'(z) - A'(z)} = 1 - A'(1).$$

$A(z)$ is the generating function of number of customers arriving for the service time of one customer, i.e.

$$\begin{aligned} A(z) &= \sum_{k=0}^{\infty} a_k z^k = \sum_{k=0}^{\infty} z^k \int_0^{\infty} \frac{(\lambda x)^k}{k!} e^{-\lambda x} dB(x) = \\ &= \int_0^{\infty} \sum_{k=0}^{\infty} \frac{(\lambda z x)^k}{k!} e^{-\lambda x} dB(x) = \int_0^{\infty} e^{-\lambda x(1-z)} dB(x) = b(\lambda(1-z)). \end{aligned}$$

Here $B(x)$ is the distribution function of service time of a customer, and

$$b(s) = \int_0^{\infty} e^{-sx} dB(x)$$

is its Laplace-Stieltjes transform. Consequently,

$$A'(1) = -\lambda b'(\lambda(1-z))|_{z=1} = -\lambda b'(0) = \lambda\tau,$$

where τ is the mean value of service time of a customer.

In the case of M/M/1 system the stability condition was the fulfilment of inequality

$$\rho = \lambda \cdot \frac{1}{\mu} < 1.$$

ρ is the product of the intensity of arrivals and the mean value of service time. In case of the M/G/1 system the situation is similar: the intensity of arrivals (λ) is multiplied by the mean value of service time (τ), so it is reasonable to use the notation $\rho = \lambda\tau$. The generating function for the number of present customers may be written in the form

$$P(z) = \frac{(1-\rho)(z-1)A(z)}{z - A(z)} = \frac{(1-\rho)(1-z)b(\lambda(1-z))}{b(\lambda(1-z)) - z}.$$

This expression is known as Pollaczek-Khinchin formula in the Russian language and as Pollaczek-Khinchin transform equation in the English language literature.

2.2. 2.2 The distribution function of busy period

As it has been mentioned earlier at the solution of the Takács integro-differential equation by means of the Laplace-Stieltjes transform one has to know the probability of free state as the function of time. From another side the busy period is an important feature of the system, it offers information about the requirements concerning the server.

The idea of busy period is very simple: the free state ends with the entry of a customer and its service begins. During this service further customers arrive, they will also be served, and for these services again new customers are generated. This process is continued and it will be terminated when during the service of the only present customer no further ones arrive. One can formulate the idea of busy period on the manner: the busy period means the service of a customer with generated by it ones.

There play important role the customers arriving for the service of first one, their services with the generated by them ones have the same structure as that of the busy period.

Let $G(x)$ be the distribution function of the busy period. The busy period consists of two parts: the service of first customer and the service of all other ones. Let y be the service time of first customer, during this time k new customers enter with probability

$$\frac{(\lambda y)^k}{k!}e^{-\lambda y}.$$

We have to serve all these customers with the generated by them ones, the structure of these services coincides with the structure of the busy period. So,

$$G(x) = \int_0^x \sum_{k=0}^{\infty} \frac{(\lambda y)^k}{k!} e^{-\lambda y} G_k(x-y) dB(y),$$

where $G_k(x)$ denotes the k -fold convolution of distribution function $G(x)$. Let

$$\Gamma(s) = \int_0^{\infty} e^{-sx} dG(x), \qquad b(s) = \int_0^{\infty} e^{-sx} dB(x),$$

then the Laplace-Stieltjes transform of distribution function $G(x)$ is

$$\begin{aligned}
\Gamma(s) &= \sum_{k=0}^{\infty} \frac{1}{k!} [\Gamma(s)]^k \int_0^{\infty} e^{-(s+\lambda)x} (\lambda x)^k dB(x) = \\
&= \sum_{k=0}^{\infty} (-1)^k \frac{\lambda^k [\Gamma(s)]^k b^{(k)}(s+\lambda)}{k!} = \\
&= b(s+\lambda-\lambda\Gamma(s)),
\end{aligned}$$

where $b^{(k)}(s)$ is the k -th derivative of the function $b(s)$, the second row is Taylor expansion of the function $b(s+\lambda-\lambda\Gamma(s))$. So, the desired transform is the solution of functional equation

$$\Gamma(s) = b(s+\lambda-\lambda\Gamma(s)).$$

We show that this functional equation has solution. Let us introduce the notation

$$s+\lambda-\lambda\Gamma(s) = x,$$

then

$$\frac{s+\lambda-x}{\lambda} = b(x).$$

Since

$$b(s) = \int_0^{\infty} e^{-sx} dB(x)$$

is a Laplace-Stieltjes transform, we have

$$b(0) = 1, \qquad b'(s) = -\int_0^{\infty} x e^{-sx} dB(x) < 0, \qquad b''(s) = \int_0^{\infty} x^2 e^{-sx} dB(x) > 0,$$

i.e. $b(s)$ is a concave, monotonically decreasing on $[0,\infty)$ function, at the point 0 takes on the value 1 and asymptotically approaches the axis x .

$$\frac{s+\lambda-x}{\lambda}$$

is the equation of a straight line, at the point 0 takes on a value greater than 1, in the case $x=s$ its value is 1, and at $x=s+\lambda$ intersects the axis x . Consequently, the straight line and the Laplace-Stieltjes transform (the left and right sides of the above equation) intersect between s and $s+\lambda$, for a fixed value of s this will be the solution, the corresponding value of function $\Gamma(s)$.

We mention in case of general distribution of service time this functional equation cannot be solved explicitly, one has to find numerically the value of $\Gamma(s)$ for all values of s . From these values one can approximate $\Gamma(s)$, its inverse transform will be the distribution function of the busy period.

As we see, to find the distribution function of busy period is rather complicated, but the mean value of its duration can easily be computed from the functional equation

$$\Gamma(s)=b(s+\lambda-\lambda\Gamma(s)).$$

By means of differentiation we obtain

$$\Gamma'(s)=b'(s+\lambda-\lambda\Gamma(s))[1-\lambda\Gamma'(s)],$$

from which

$$\Gamma'(s)=\frac{b'(s+\lambda-\lambda\Gamma(s))}{1+\lambda b'(s+\lambda-\lambda\Gamma(s))}$$

and

$$-\Gamma'(0)=\frac{\tau}{1-\lambda\tau}=\frac{\tau}{1-\rho},$$

where $\tau=\int\limits_0^\infty x dB(x)$ is the mean value of service time for a customer.

The solution of Takács integro-differential equation requires the knowledge of probability $F(t,0)$. One can find it by using the Laplace-Stieljes transform of the busy periods distribution function $\Gamma(s)$.

If we consider the moments of transitions from busy state to free state, they will be the regeneration points of a renewal process. The interval between two such points is the sum of two random variables: that of an exponentially distributed one and a busy period. At the moment t the system is in free state if 1. till t no customer entered the system; 2. at moment τ after a busy period the system became free and for $t-\tau$ no new customers arrived. The last renewal point may be arbitrary, so the event the system is free at t is the sum of mutually exclusive events, i.e.

$$\begin{aligned}F(t,0) &= e^{-\lambda t}+\sum_{n=1}^{\infty}\int\limits_0^t e^{-\lambda(t-\tau)}dF^{(n)}(\tau)= \\ &= e^{-\lambda t}+\int\limits_0^t e^{-\lambda(t-\tau)}d\left(\sum_{n=1}^{\infty}F^{(n)}(\tau)\right)= \\ &= e^{-\lambda t}+\int\limits_0^t e^{-\lambda(t-\tau)}dH(\tau),\end{aligned}$$

where $H(t)=\sum_{n=1}^{\infty}F^{(n)}(t)$ is the renewal function of the process. The Laplace-Stieltjes transform of distribution function of the interval between two renewal points is

$$\frac{\lambda}{\lambda+s}\Gamma(s)$$

(the distribution function is the convolution of an exponential distribution and a busy period), the Laplace-Stieltjes transform of the renewal function is

$$h(s)=\int_0^{\infty}e^{-sx}dH(x)=\sum_{n=1}^{\infty}\left[\frac{\lambda\Gamma(s)}{\lambda+s}\right]^n=\frac{\frac{\lambda\Gamma s}{\lambda+s}}{1-\frac{\lambda\Gamma(s)}{\lambda+s}}=\frac{\lambda\Gamma(s)}{s+\lambda-\lambda\Gamma(s)}.$$

Since the probability of free state is

$$F(t,0)=e^{-\lambda t}+\int_0^t e^{-\lambda(t-\tau)}dH(\tau),$$

by using the expression for $h(s)$ we obtain the Laplace transform of probability of free state

$$\int_0^{\infty}e^{-st}F(t,0)dt=\frac{1}{s+\lambda}+\frac{1}{s+\lambda}\frac{\lambda\Gamma(s)}{s+\lambda-\lambda\Gamma(s)}=\frac{1}{s+\lambda-\lambda\Gamma(s)}.$$

2.3. 2.3 Another approach to the Pollaczek-Khinchin formula

Earlier we derived the generating function of number of customers in the M/G/1 system by means of the embedded Markov chain. From this generating function on the most natural way these probabilities may be obtained by means of differentiation. Even in the 80s there were ideas to look for simpler methods, in case of rational Laplace-Stieltjes transform of service time there were certain successes. Later one applied the FFT method, it gave an effective tool to the series expansion.

Some systems may effectively be described by means of regenerative processes. It means that during its functioning the system is renewed and after these regeneration points its stochastic behaviour will be the same. A simple example is a technical equipment which after each maintenance is functioning as a "new" one. The interval between two successive regeneration points is called a "regenerative cycle", their lengths are independent identically distributed random variables. Inside a cycle the system may stay in different states. On the basis of results concerning the regenerative processes one can find the stationary probabilities (see Tijms (1994), Lakatos (2013)) as the relations of the mean values of times spent in different states to the mean value of the regenerative cycle. Now we use this approach to the M/G/1 system.

Let us introduce the notations:

ζ - the mean value of length of the busy period;

ζ_i - the mean value of time spent above the i -th level for a busy period;

ξ_i - the mean value of time on the i -th level for a busy period.

Theorem. Let us consider the M/G/1 system. The arrival process is Poisson with parameter λ , the distribution function of service time is $B(x)$. If the mean value of service time of a customer τ is finite, $\lambda\tau<1$, then there exists an equilibrium distribution. This distribution is given by the fractions $p_i=\xi_i/\zeta \quad (i=0,1,2,\dots)$, where ζ is the mean value of the busy period, ξ_i is the mean value of time spent on the i -th level for a busy period.

The mean value of busy period was earlier computed, it remains to find the mean values times spent in different states.

Lemma. In the M/G/1 system

$$\xi_0 = \tau, \qquad \xi_1 = \frac{1-a_0}{a_0}\tau, \qquad \xi_2 = \frac{1-a_0-a_1}{a_0}(\xi_0+\xi_1)$$

and the values $\xi_k \ (k \geq 3)$ satisfy the recurrence relation

$$\xi_k = \sum_{i=1}^{k-2} \frac{1-a_0-a_1-\ldots-a_i}{a_0}\xi_{k-i} + \frac{1-a_0-a_1-\ldots-a_{k-1}}{a_0}(\xi_0+\xi_1).$$

At the derivation of the Pollaczek-Hincsin transform equation the system was characterized by the number of customers remaining in the system after having served them. For the future computations it will be more convenient to consider the number of customers at the beginning of services. So, we will have to distinguish the states and the number of customers, the difference between them will be clarified below.

If at the beginning of service of a customer it is the only present one, the number of customers is obviously 1. The state is determined by the fact how many customers remain there after the service. If at least two customers arrive then this service belongs to a higher level; if one customer enters then to the state 1; if no customer arrives then to the state 0, this situation takes place at the service of last customer in the busy period. If we examine the relation between the states and the number of customers then in the first case the number of present customers is decreased by one, but we get back this value returning to the first level since on the second level at the service of last customer two ones are present, but this service already belongs to the state 1. In the case of first level the situation is the same excluding the service of last customer. In the case of higher levels we get one customer coming to this level, and lose one going one level down. Consequently, for the number of customers we get the same value as in the case of states.

Similarly to the investigation by means of the embedded Markov chain we will consider the state of system at moments just after having served the individual customers, i.e. when it already left the system. Let j customers be present. After the actual service with probability a_1 we remain on the same level (one new customer entered and one service was completed), with probability $1-a_1$ we leave the level; more exactly with probability $\frac{a_0}{1-a_1}$ we come to the $j-1$-st and with probability $\frac{1-a_0-a_1}{1-a_1}$ to a higher level.

During a period when only one customer is in the system on average we serve

$$\sum_{k=1}^{\infty} k a_1^{k-1}(1-a_1) = \frac{1}{1-a_1}$$

ones (in $k-1$ cases one new customer enters, in the last case we come to either free state or to a higher level).

Now we take a period during which we are above the first level. The average number of customers served for it is

$$\sum_{k=2}^{\infty} \frac{a_k}{1-a_0-a_1}(k-1)\frac{1}{1-\rho} = \frac{1}{(1-\rho)(1-a_0-a_1)}[\rho - a_1 - (1-a_0-a_1)] =$$

$$= \frac{\rho-1+a_0}{(1-\rho)(1-a_0-a_1)},$$

where we used the equalities

$$\rho = \sum_{k=1}^{\infty} k a_k \qquad \text{and} \qquad \sum_{k=0}^{\infty} a_k = 1.$$

(From the first level we come to a higher one with probability $1-a_0-a_1$, on this condition with probability $\frac{a_k}{1-a_0-a_1}$ to the k-th one. In order to return to the first level we have to serve $k-1$ of them, but with the generated by them ones. This process starts with the service of one customer and is terminated when the last customer belonging to it leaves the system, i.e. its structure corresponds to the structure of the busy period, the mean number of served customers is $(k-1)(1-\rho)^{-1}$.)

During the busy period there change the intervals on and above the first level. The intervals on the first level may be terminated on two different ways: either no customer arrives (it means the end of busy period), or more than one customers enter. For a busy period we have $0,\ 1,\ 2,\dots,\ k,\dots$ intervals above the first level with probabilities

$$\frac{a_0}{1-a_1},\quad \frac{1-a_0-a_1}{1-a_1}\frac{a_0}{1-a_1},\dots,\frac{(1-a_0-a_1)^k}{(1-a_1)^k}\frac{a_0}{1-a_1},\dots$$

By using these probabilities the mean values of time spent on and above the first level are

$$\sum_{k=1}^{\infty}k\frac{(1-a_0-a_1)^{k-1}}{(1-a_1)^{k-1}}\frac{a_0}{1-a_1}\frac{\tau}{1-a_1}=\frac{\tau}{a_0},$$

$$\sum_{k=1}^{\infty}k\frac{(1-a_0-a_1)^k}{(1-a_1)^k}\frac{a_0}{1-a_1}\frac{\rho-1+a_0}{(1-\rho)(1-a_0-a_1)}\tau=\frac{\rho-1+a_0}{a_0(1-\rho)}\tau.$$

The sum of two values is

$$\frac{\tau}{a_0}+\frac{\rho-1+a_0}{a_0(1-\rho)}\tau=\frac{\tau}{1-\rho},$$

i.e. it gives the mean value of busy period.

Our above computations were based on the number of present customers, so the mean value of time on the first level means the sum of times in states 0 and 1, i.e. $\xi_0+\xi_1$. To the zero state in a busy period belongs only one customer, consequently $\xi_0=\tau$ and

$$\xi_1=\frac{\tau}{a_0}-\tau=\frac{1-a_0}{a_0}\tau.$$

We derive the mean value of time spent above the k -th level. First we consider the second level. We have two possibilities:

1. from the first level we come to the second one;

2. from the first level we come at least to the third level.

When we come from the first level to the second one we are in the same situation as in case of the first level: serving a certain number of customers on the second level we come to the first one or above the second one. In the first case the sojourns on and above the second level change, and spending on average ζ_1 time above the second level we come to the first one. In the second case from the first level we jump above the second one, the return time to the second level is

$$\sum_{k=3}^{\infty}\frac{a_k}{1-a_0-a_1-a_2}(k-2)\frac{\tau}{1-\rho}=\frac{\rho-2+2a_0+a_1}{(1-\rho)(1-a_0-a_1-a_2)}\tau=\varepsilon_2.$$

Now we are in the previous situation and spend ζ_1 time above the second level. The probabilities of two possibilities are

$$\frac{a_2}{1-a_0-a_1} \quad \text{and} \quad \frac{1-a_0-a_1-a_2}{1-a_0-a_1},$$

so the average sojourn time above the second level for a period beginning and ending on the first level is

$$\frac{a_2}{1-a_0-a_1}\zeta_1 + \frac{1-a_0-a_1-a_2}{1-a_0-a_1}(\zeta_1+\varepsilon_2) = \zeta_1 + \varepsilon_2',$$

where

$$\varepsilon_2' = \frac{\rho-2+2a_0+a_1}{(1-\rho)(1-a_0-a_1)}\tau.$$

During a busy period we will have i such intervals with probability $\frac{(1-a_0-a_1)^i}{(1-a_1)^i}\frac{a_0}{1-a_1}$, so

$$\zeta_2 = \sum_{i=1}^{\infty} i\frac{(1-a_0-a_1)^i}{(1-a_1)^i}\frac{a_0}{1-a_1}(\zeta_1+\varepsilon_2') = \frac{1-a_0-a_1}{a_0}\zeta_1 + \frac{1-a_0-a_1-a_2}{a_0}\varepsilon_2.$$

The mean value of time spent on the second level is obtained as the difference of mean values of times spent above the first and second levels. Since

$$\frac{1-a_0-a_1-a_2}{a_0}\varepsilon_2 = \frac{\rho-2+2a_0+a_1}{a_0(1-\rho)}\tau = \frac{\rho-1+a_0}{a_0(1-\rho)}\tau - \frac{1-a_0-a_1}{a_0(1-\rho)}\tau =$$

$$= \zeta_1 - \frac{1-a_0-a_1}{a_0}\frac{\tau}{1-\rho},$$

consequently

$$\xi_2 = \zeta_1 - \zeta_2 = \zeta_1 - \frac{1-a_0-a_1}{a_0}\zeta_1 - \frac{1-a_0-a_1-a_2}{a_0}\varepsilon_2 =$$

$$= \zeta_1 - \frac{1-a_0-a_1}{a_0}\zeta_1 - \zeta_1 + \frac{1-a_0-a_1}{a_0}\frac{\tau}{1-\rho} =$$

$$= \frac{1-a_0-a_1}{a_0}(\zeta-\zeta_1) = \frac{1-a_0-a_1}{a_0}(\xi_0+\xi_1).$$

We compute the mean value of time spent on the third level. From the first level one can jump to the second, the third and above the third level. In the three cases the mean values of times above the third level are

- ζ_2,

- $\zeta_1+\zeta_2$,

- $\varepsilon_3+\zeta_1+\zeta_2$.

In the first case from the first level we come to the second one, the mean value of time spent above the third level coincides with the case of second level from the viewpoint of first level, i.e. the desired mean value is ζ_2. In the second case the service begins on the third level, spending ζ_1 time above the the third level we come to the second level and will be in the previous situation. The corresponding mean value is $\zeta_1+\zeta_2$. In the third case the service starts above the third level, let it be k, having served $k-3$ customers and the generated by them ones we reach the third level and the second case takes place. The mean value of time to return to the third level is

$$\varepsilon_3=\sum_{k=4}^{\infty}\frac{a_k}{1-a_0-a_1-a_2-a_3}(k-3)\frac{\tau}{1-\rho}=\frac{\rho-3+3a_0+2a_1+a_2}{1-a_0-a_1-a_2-a_3}\frac{\tau}{1-\rho}.$$

The probabilities of three cases are

$$\frac{a_0}{1-a_0-a_1},\qquad \frac{a_3}{1-a_0-a_1}\quad\text{and}\quad \frac{1-a_0-a_1-a_2-a_3}{1-a_0-a_1},$$

so the mean value of time spent above the third level for an interval beginning and ending on the first level is

$$\frac{a_2}{1-a_0-a_1}\zeta_2+\frac{a_3}{1-a_0-a_1}(\zeta_1+\zeta_2)+\frac{1-a_0-a_1-a_2-a_3}{1-a_0-a_1}(\varepsilon_3+\zeta_1+\zeta_2)=$$

$$=\zeta_2+\frac{1-a_0-a_1-a_2}{1-a_0-a_1}\zeta_1+\frac{1-a_0-a_1-a_2-a_3}{1-a_0-a_1}\varepsilon_3.$$

For a busy period we have i intervals above the first level with probability $\frac{(1-a_0-a_1)^i}{(1-a_1)^i}\frac{a_0}{1-a_1}$, the mean value of time above the third level is

$$\zeta_3=\sum_{i=1}^{\infty}i\frac{(1-a_0-a_1)^i}{(1-a_1)^i}\frac{a_0}{1-a_1}\left[\zeta_2+\frac{1-a_0-a_1-a_2}{1-a_0-a_1}\zeta_1+\frac{1-a_0-a_1-a_2-a_3}{1-a_0-a_1}\varepsilon_3\right]=$$

$$=\frac{1-a_0-a_1}{a_0}\zeta_2+\frac{1-a_0-a_1-a_2}{a_0}\zeta_1+\frac{\rho-3+3a_0+2a_1+a_2}{a_0}\frac{\tau}{1-\rho}.$$

We find the mean value of time on the third level as the difference of mean values above the the second and third levels, it is

$$\xi_3=\zeta_2-\zeta_3,$$

where

$$\zeta_2=\frac{1-a_0-a_1}{a_0}\zeta_1+\frac{\rho-2+2a_0+a_1}{a_0}\frac{\tau}{1-\rho},$$

$$\zeta_3=\frac{1-a_0-a_1}{a_0}\zeta_2+\frac{1-a_0-a_1-a_2}{a_0}\zeta_1+\frac{\rho-3+3a_0+2a_1+a_2}{a_0}\frac{\tau}{1-\rho}.$$

Since

$$\frac{\rho-2+2a_0+a_1}{a_0}\frac{\tau}{1-\rho}-\frac{\rho-3+3a_0+2a_1+a_2}{a_0}\frac{\tau}{1-\rho}=$$

$$=\frac{1-a_0-a_1-a_2}{a_0}\frac{\tau}{1-\rho}=\frac{1-a_0-a_1-a_2}{a_0}\zeta,$$

so

$$\xi_3=\frac{1-a_0-a_1}{a_0}(\zeta_1-\zeta_2)-\frac{1-a_0-a_1-a_2}{a_0}\zeta_1+\frac{1-a_0-a_1-a_2}{a_0}\zeta=$$

$$= \frac{1 - a_0 - a_1}{a_0}\xi_2 + \frac{1 - a_0 - a_1 - a_2}{a_0}(\xi_0 + \xi_1),$$

i.e. in the case $k = 3$ the statement of the lemma is true.

Now we consider the k -th level and compute the mean value of time spent on it.

From the first level we can come to the second, third,..., k -th level and to a level above k . For ζ_k we have the following possibilities:

$$\begin{array}{ll} \zeta_k : & \zeta_{k-1} \\ & \zeta_{k-2} + \zeta_{k-1} \\ & \ldots\ldots\ldots\ldots\ldots \\ & \zeta_{k-i} + \zeta_{k-i+1} + \ldots + \zeta_{k-1} \\ & \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ & \zeta_1 + \zeta_2 + \ldots + \zeta_{k-1} + \varepsilon_k \end{array}$$

The first possibility is the second level. We are in the situation when we examine the time spent above the $k-1$-st level from the viewpoint of first level, so the mean value is ζ_{k-1}.

In the case of third level first we have a period beginning with three and ending two present customers, this corresponds to the time spent above the $k-2$-nd level from the viewpoint of first one (the mean value is ζ_{k-2}). After it we are in the earlier situation. Consequently, the mean value of time spent above the k -th level for an interval beginning on the third and ending on the first level is $\zeta_{k-2} + \zeta_{k-1}$.

Let us consider the last case. From the first level we jump above the k -th one, let it be the i -th. The mean value of time to return to the k -th level is

$$\sum_{i=k+1}^{\infty} \frac{a_i}{1 - a_0 - a_1 - \ldots - a_k}(i-k)\frac{\tau}{1-\rho} =$$

$$= \frac{\tau}{(1-\rho)(1 - a_0 - a_1 - \ldots - a_k)}\{\rho - a_1 - 2a_2 - \ldots - ka_k - k(1 - a_0 - a_1 - \ldots - a_k)\}$$

$$= \frac{\rho - k + ka_0 + (k-1)a_1 + \ldots + 2a_{k-2} + a_{k-1}}{(1-\rho)(1 - a_0 - a_1 - \ldots - a_k)}\tau = \varepsilon_k.$$

After that we stay on the k -th level, spending on average ζ_1 above the k -th we come to the $k-1$-st, spending ζ_2 above the k -th we come to the $k-2$-nd,..., and, finally, starting from the second level and spending ζ_{k-1} time above the k -th level we reach the first level. So, in the last case the mean value of time spent above the k -th level is $\zeta_1 + \zeta_2 + \ldots + \zeta_{k-2} + \zeta_{k-1} + \varepsilon_k$. The probability of first case is $\frac{a_2}{1 - a_0 - a_1}$, the probability of second case is $\frac{a_3}{1 - a_0 - a_1}$,..., the probability of last case is $\frac{1 - a_0 - a_1 - \ldots - a_k}{1 - a_0 - a_1}$. The mean value of time spent above the k -th level for a period beginning and ending on the first level is

$$\frac{a_2}{1 - a_0 - a_1}\zeta_{k-1} + \frac{a_3}{1 - a_0 - a_1}(\zeta_{k-2} + \zeta_{k-1}) + \ldots + \frac{a_k}{1 - a_0 - a_1}(\zeta_1 + \zeta_2 + \ldots + \zeta_{k-1}) +$$

$$+ \frac{1 - a_0 - a_1 - \ldots - a_k}{1 - a_0 - a_1}(\zeta_1 + \zeta_2 + \ldots + \zeta_{k-1} + \varepsilon_k) =$$

$$= \zeta_{k-1} + \frac{1-a_0-a_1-a_2}{1-a_0-a_1}\zeta_{k-2} + \ldots +$$

$$+ \frac{1-a_0-a_1-\ldots-a_{k-1}}{1-a_0-a_1}\zeta_1 + \frac{1-a_0-a_1-\ldots-a_k}{1-a_0-a_1}\varepsilon_k.$$

For a busy period we have i such intervals with probability $\frac{(1-a_0-a_1)^i}{(1-a_1)^i}\frac{a_0}{1-a_1}$, the mean value of time spent above the k -th level for a busy period is

$$\zeta_k = \sum_{i=0}^{\infty} i \frac{(1-a_0-a_1)^i}{(1-a_1)^i}\frac{a_0}{1-a_1}\left\{\zeta_{k-1} + \frac{1-a_0-a_1-a_2}{1-a_0-a_1}\zeta_{k-2} + \ldots + \right.$$

$$\left. + \frac{1-a_0-a_1-\ldots-a_{k-1}}{1-a_0-a_1}\zeta_1 + \frac{1-a_0-a_1-\ldots-a_k}{1-a_0-a_1}\varepsilon_k\right\} =$$

$$= \frac{1-a_0-a_1}{a_0}\zeta_{k-1} + \frac{1-a_0-a_1-a_2}{a_0}\zeta_{k-2} + \ldots +$$

$$+ \frac{1-a_0-a_1-\ldots-a_{k-1}}{a_0}\zeta_1 + \frac{1-a_0-a_1-\ldots-a_k}{a_0}\varepsilon_k =$$

$$= \sum_{i=1}^{k-1} \frac{1-a_0-a_1-\ldots-a_i}{a_0}\zeta_{k-i} + \frac{1-a_0-a_1-\ldots-a_k}{a_0}\varepsilon_k.$$

The similar value for ζ_{k-1} is

$$\zeta_{k-1} = \sum_{i=1}^{k-2} \frac{1-a_0-\ldots-a_i}{a_0}\zeta_{k-1-i} + \frac{1-a_0-\ldots-a_{k-1}}{a_0}\varepsilon_{k-1}.$$

The difference of these two values gives the mean value of time spent on the k -th level

$$\zeta_{k-1} - \zeta_k =$$

$$= \frac{1-a_0-a_1}{a_0}\zeta_{k-2} + \frac{1-a_0-a_1-a_2}{a_0}\zeta_{k-3} + \ldots + \frac{1-a_0-\ldots-a_{k-2}}{a_0}\zeta_1 +$$

$$+ \frac{\rho-(k-1)+(k-1)a_0+\ldots+2a_{k-3}+a_{k-2}}{a_0}\frac{\tau}{1-\rho} -$$

$$- \frac{1-a_0-a_1}{a_0}\zeta_{k-1} - \frac{1-a_0-a_1-a_2}{a_0}\zeta_{k-2} - \ldots - \frac{1-a_0-a_1-\ldots-a_{k-1}}{a_0}\zeta_1 -$$

$$- \frac{\rho-k+ka_0+(k-1)a_1+\ldots+2a_{k-2}+a_{k-1}}{a_0}\frac{\tau}{1-\rho}.$$

Since

$$\frac{\rho-(k-1)+(k-1)a_0+\ldots+2a_{k-3}+a_{k-2}}{a_0} -$$

$$-\frac{\rho-k+ka_0+(k-1)a_1+\ldots+2a_{k-2}+a_{k-1}}{a_0}=$$

$$=\frac{1-a_0-a_1-\ldots-a_{k-1}}{a_0},$$

so

$$\xi_k=\zeta_{k-1}-\zeta_k=$$

$$=\frac{1-a_0-a_1}{a_0}(\zeta_{k-2}-\zeta_{k-1})+\frac{1-a_0-a_1-a_2}{a_0}(\zeta_{k-3}-\zeta_{k-2})+...+$$

$$+\frac{1-a_0-\ldots-a_{k-2}}{a_0}(\zeta_1-\zeta_2)-\frac{1-a_0-a_1-\ldots-a_{k-1}}{a_0}\zeta_1+\frac{1-a_0-a_1-\ldots-a_{k-1}}{a_0}\frac{\tau}{1-\rho}$$

$$=\sum_{i=1}^{k-2}\frac{1-a_0-\ldots-a_i}{a_0}\xi_{k-i}+\frac{1-a_0-a_1-\ldots-a_{k-1}}{a_0}(\xi_0+\xi_1).$$

The lemma is proved, the desired ergodic probabilities may be computed according to our theorem. Now we show the Pollaczek-Khinchin transform equation may be derived from these mean values.

Theorem. The generating function of number of customers in the M/G/1 system

$$P(z)=\frac{(1-\rho)(1-z)b(\lambda(1-z))}{b(\lambda(1-z))-z},$$

may be derived on the basis of results for the regenerative processes from the mean values

$$\xi_0=\tau,\qquad \xi_1=\frac{1-a_0}{a_0}\tau,\qquad \xi_2=\frac{1-a_0-a_1}{a_0}(\xi_0+\xi_1),$$

$$\xi_k=\sum_{i=1}^{k-2}\frac{1-a_0-a_1-\ldots-a_i}{a_0}\xi_{k-i}+\frac{1-a_0-a_1-\ldots-a_{k-1}}{a_0}(\xi_0+\xi_1).$$

Remark. The above generating function for the number of customers in the M/G/1 system is a classical result, it is the so-called Pollaczek-Khinchin transform equation. According to the definition of embedded Markov chain the busy period means the regenerative cycle, the zero state is the service of last customer in the busy period.

Let us write the formulas for the first values of ξ_i:

$$
\begin{aligned}
\xi_0 &= \xi_0, \\
\xi_1 &= \frac{1-a_0}{a_0}\xi_0, \\
\xi_2 &= \frac{1-a_0-a_1}{a_0}\xi_1 + \frac{1-a_0-a_1}{a_0}\xi_0, \\
\xi_3 &= \frac{1-a_0-a_1}{a_0}\xi_2 + \frac{1-a_0-a_1-a_2}{a_0}\xi_1 + \frac{1-a_0-a_1-a_2}{a_0}\xi_0, \\
\xi_4 &= \frac{1-a_0-a_1}{a_0}\xi_3 + \frac{1-a_0-a_1-a_2}{a_0}\xi_2 + \\
&\quad + \frac{1-a_0-a_1-a_2-a_3}{a_0}\xi_1 + \frac{1-a_0-a_1-a_2-a_3}{a_0}\xi_0, \\
\xi_5 &= \frac{1-a_0-a_1}{a_0}\xi_4 + \frac{1-a_0-a_1}{a_0}\xi_4 + \\
&\quad + \frac{1-a_0-a_1-a_2}{a_0}\xi_3 + \frac{1-a_0-a_1-a_2-a_3}{a_0}\xi_2 + \\
&\quad + \frac{1-a_0-a_1-a_2-a_3-a_4}{a_0}\xi_1 + \frac{1-a_0-a_1-a_2-a_3-a_4}{a_0}\xi_0, \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned}
$$

Multiply the expression for ξ_i by z^i and sum up from the third row (the formula for ξ_2) excluding the last term (containing ξ_0). Then we obtain

$$
\begin{aligned}
&\frac{1-a_0-a_1}{a_0}z(\xi_1 z + \xi_2 z^2 + \ldots) + \frac{1-a_0-a_1-a_2}{a_0}z^2(\xi_1 z + \xi_2 z^2 + \ldots) + \\
&\quad + \frac{1-a_0-a_1-a_2-a_3}{a_0}z^3(\xi_1 z + \xi_2 z^2 + \ldots) + \ldots = \\
&= \left(\sum_{i=1}^{\infty}\xi_i z^i\right)\left\{\frac{1-a_0-a_1}{a_0}z + \frac{1-a_0-a_1-a_2}{a_0}z^2 + \right. \\
&\quad \left. + \frac{1-a_0-a_1-a_2-a_3}{a_0}z^3 + \ldots\right\} = \\
&= \left(\sum_{i=1}^{\infty}\xi_i z^i\right)\frac{1}{a_0}\left\{\frac{z}{1-z} - \frac{a_0 z}{1-z} - \frac{a_1 z}{1-z} - \frac{a_2 z^2}{1-z} - \frac{a_3 z^3}{1-z} - \ldots\right\} = \\
&= \left(\sum_{i=1}^{\infty}\xi_i z^i\right)\frac{1}{a_0(1-z)}\left\{z(1-a_0) - [A(z) - a_0]\right\} = \\
&= \left[\overline{P}(z) - \xi_0\right]\frac{1}{a_0(1-z)}\left\{z(1-a_0) - [A(z) - a_0]\right\},
\end{aligned}
$$

where $\overline{P}(z) = \sum_{i=0}^{\infty}\xi_i z^i$. Similarly, for the terms containing ξ_0 we get

$$
\begin{aligned}
& \xi_0 z \sum_{i=1}^{\infty} \frac{1-a_0-a_1-\ldots-a_i}{a_0} z^i = \\
= \; & \frac{\xi_0 z}{a_0}\left\{(1-a_0-a_1)z+(1-a_0-a_1-a_2)z^2+(1-a_0-a_1-a_2-a_3)z^3+\ldots\right\} = \\
= \; & \frac{\xi_0 z}{a_0}\left\{\frac{z}{1-z}-\frac{a_0}{1-z}-\frac{a_1 z}{1-z}-\frac{a_2 z^2}{1-z}-\frac{a_3 z^3}{1-z}-\ldots\right\} = \\
= \; & \xi_0 z \frac{1}{a_0(1-z)}\left\{z(1-a_0)-[A(z)-a_0]\right\}.
\end{aligned}
$$

Adding the above two expressions, the first row and the second row multiplied by z , we obtain

$$
\overline{P}(z)=[\overline{P}(z)-\xi_0]\frac{1}{a_0(1-z)}\{z(1-a_0)-[A(z)-a_0]\}+
$$

$$
+\xi_0 z\frac{1}{a_0(1-z)}\{z(1-a_0)-[A(z)-a_0]\}+\xi_0+\frac{1-a_0}{a_0}\xi_0 z,
$$

or

$$
\begin{aligned}
(a_0-a_0 z)\overline{P}(z) \;=\; & [\overline{P}(z)-\xi_0][z-a_0 z-A(z)+a_0]+ \\
& +\xi_0 z[z-a_0 z-A(z)+a_0]+\xi_0(a_0-a_0 z)+\xi_0 z(1-a_0)(1-z),
\end{aligned}
$$

from which

$$
\overline{P}(z)=\frac{(1-z)A(z)}{A(z)-z}\xi_0.
$$

Dividing this formula by the mean value of busy period

$$
\frac{\tau}{1-\rho}
$$

and taking into account that $\xi_0=\tau$, we obtain

$$
P(z)=\frac{(1-\rho)(1-z)A(z)}{A(z)-z}=\frac{(1-\rho)(1-z)b(\lambda(1-z))}{b(\lambda(1-z))}.
$$

3. 3 Cyclic-waiting systems

According to the Kendall notation a queueing system is characterized by the interarrival and service times, the number of servers and the waiting room. This notation does not include the service discipline which plays key role, too. It determines the order of service, these rules may be rather simple (first-come-first-served, last-come-first-served, random, etc.) or more complex depending on the waiting time, number of present customers or priorities and so on. The analysis of queueing system with simple probabilistic characteristics may be rather complicated because of the service discipline.

As we know the queueing theory at the very beginning was connected with the telephone systems, but there was not raised the natural question: what happens with the refused calls? In Erlangs time this question was solved on a very simple way. They are lost and their repetitions can be considered as new calls. Later, with the development of theory, a more sophisticated approach appeared, the system distinguished the primary (appearing for the first time) and secondary (at least once refused) calls. This separation led to the so-called retrial systems.

We propose to consider a single-server queueing system, where an entering customer may be accepted for service either at the moment of arrival or at moments differing from it by the multiples of a given so-called cycle time. In order to illustrate the problem we give two practical examples.

1. Airplanes arrive at the airport in optimal position for landing. If there is no queue and the previous one is far enough, they start the landing process. If the distance is too small or there are some waiting ones, they start circling. The next request for service may be put when the airplane arrives at the starting geometrical point and this procedure is repeated.

2. Optical signals enter a node and they should be transmitted according to the FCFS rule. This information cannot be stored, if the immediate transmission is impossible it is sent to a delay line and returns to the node after having passed it. Clearly, the signal can be transmitted from the node at the time of its arrival or at the time that differs from it by a time multiple of time necessary to pass the delay line.

The queueing systems may be considered from the viewpoints of the system and the individual customers. From the viewpoint of the system the number of present customers is important, from the viewpoint of individual customers the waiting time plays essential role.

3.1. 3.1 Number of customers

3.1.1. 3.1.1 The continuous time case

Let us consider a queueing system with Poisson arrivals. If the server is free, the service starts immediately. Otherwise, the entering customer joins the waiting queue from which it is taken in the order of arrival and the service can start at a moment differing from the arrival time by a multiple of cycle time T. The system is characterized by N_{t_n}, the number of customers in the system at the instant just before starting the service of n-th customer. We show that these quantities form a Markov chain.

Let t_n be the moment of the beginning of service of n-th customer. The number of customers in the system at t_{n+1} is

$$N_{t_{n+1}} = \begin{cases} \Delta_n, & \text{if } N_{t_n} = 0, \\ N_{t_n} - 1 + \Delta_n, & \text{if } N_{t_n} > 0, \end{cases}$$

where Δ_n is the number of customers arriving at the system for $[t_n, t_{n+1})$. We show that Δ_n are independent random variables.

First, let us consider the intervals between two consecutive moments when we start the services of customers. Let ξ_i and η_i ($i = 1, 2, \dots$) be two sequences of independent random variables, independent of one another. ξ_i denotes the interarrival time between the i-th and $i+1$-st customers, it is exponentially distributed with parameter λ; η_i is the service time of i-th customer (in our case, it is exponentially distributed with parameter μ).

Assume that there is only one customer in the system at the beginning of service. If the interarrival time is longer than the service time ($\xi_i > \eta_i$) then the service i-th customer is completed, and after a free period enters the $i+1$-st one. In the inverse case, the next customer appears during the service, after the moment of entry we have to consider intervals of lengths T, and at the first such moment when the server is free we begin the service of the new customer. We are interested in the interval between the starting moments, it obviously will be a certain function of ξ_i and η_i, i.e. $f_1(\xi_i, \eta_i)$.

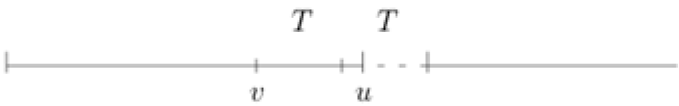
If at the beginning of service of a customer the next one is already present in the system the time interval till the beginning of its service is determined as follows. The service time of first customer is divided into intervals of length T (the last one usually is not full). Since the times of beginning of services of both customers differ from the moments of their arrivals by the multiples of T, each interval of length T will have one point where the service of second customer theoretically can be started. Actually, it begins at the first possible instant after completion of first customer; so the necessary for us time will be determined by the relation between the service time of first customer and their interarrival time, i.e. it is a certain function of random variables ξ_i and η_i, $f_2(\xi_i, \eta_i)$.

The time intervals on which the arriving customers are observed are only the functions of independent random variables ξ_i and η_i, and so they are also independent. Since the incoming flow is Poisson, the number of arriving customers for such intervals Δ_i are also independent random variables; therefore N_{t_i} form a Markov chain.

We find the transition probabilities for this chain. We distinguish two cases: at moment when the service of a customer begins the next one is present or not.

Let us consider the second possibility, it appears at states zero and one. Assume that the service time of first customer is equal to u, the second one appears v time after beginning its service.

1mm



The probability of event $\{u - v < t\}$ is equal to

$$\begin{aligned} P(t) &= P\{u - v < t\} = \\ &= \int_0^t \int_0^u \lambda e^{-\lambda v} \mu e^{-\mu u} \mathrm{d}v \mathrm{d}u + \int_t^\infty \int_{u-t}^u \lambda e^{-\lambda v} \mu e^{-\mu u} \mathrm{d}v \mathrm{d}u = \\ &= \frac{\lambda}{\lambda + \mu} \left(1 - e^{-\mu t}\right). \end{aligned} \tag{2}$$

The duration of period from the entry of second customer till the beginning of its service is equal to

$$\left(I\left(\frac{u - v}{T}\right) + 1\right) T,$$

where $I(x)$ denotes the integer part of x. This formula is valid almost everywhere, except for the multiples of cycle time T. We are interested in the number of customers arriving during this period. According to (2) the duration of this period is equal to iT with probability

$$\frac{\lambda}{\lambda + \mu} \left(e^{-\mu(i-1)T} - e^{-\mu i T}\right)$$

and the generating function of customers appearing during this time is

$$\begin{aligned} &\frac{\lambda}{\lambda + \mu} \sum_{k=0}^\infty \sum_{i=1}^\infty \left(e^{-\mu(i-1)T} - e^{-\mu i T}\right) \frac{(\lambda i T z)^k}{k!} e^{-\lambda i T} = \\ &= \frac{\lambda}{\lambda + \mu} \sum_{i=1}^\infty \left(e^{-\mu(i-1)T} - e^{-\mu i T}\right) e^{-\lambda i T(1-z)} = \frac{\lambda}{\lambda + \mu} \frac{e^{-\lambda(1-z)T}(1 - e^{-\mu T})}{1 - e^{-[\lambda(1-z)+\mu]T}}. \end{aligned}$$

The last formula was obtained on condition that during the service time one customer obligatorily appears, so the desired generating function will be

$$A(z) = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} z \frac{(1-e^{-\mu T})e^{-\lambda T(1-z)}}{1-e^{-[\lambda(1-z)+\mu]T}}, \tag{4}$$

where $\frac{\mu}{\lambda+\mu} = \int\limits_0^\infty e^{-\lambda x}\mu e^{-\mu x}dx$ is the probability of event that during the service time of a customer another one does not appear at all.

Now we determine the transition probabilities for all other states. In this case at moment when the service of the first customer begins the second one is already present, too. Let $x = u - I(\frac{u}{T})T$ and y mean the deviation of interarrival times mod T. (Consider the series of cycles starting from the entry of first customer and take the one during which the second customer enters. y means the difference between the beginning of this cycle and the arrival moment of the second customer, it obviously is equal to $x = v - I(\frac{v}{T})$.) It can easily be seen that y has truncated exponential distribution with function

$$\frac{1-e^{-\lambda y}}{1-e^{-\lambda T}}.$$

The duration of period between the starting moments of services of two consecutive customers is

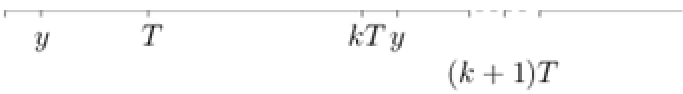
$$I\left(u/T\right)T + y \quad \text{if} \quad x \le y \qquad \text{and} \qquad \left(I\left(u/T\right)+1\right)T + y \quad \text{if} \quad x > y.$$

$x \le y$:

1.5mm

$x > y$:

1.5mm



$x = u - I(u)T$ - *mod* T service time; y - *mod* T interarrival time

We are interested in the number of customers arriving during these periods, there appear k ones with probabilities

$$\frac{\left(\lambda\left\{I\left(u/T\right)T+y\right\}\right)^k}{k!}\exp\left(-\lambda\left\{I\left(u/T\right)T+y\right\}\right) \tag{5}$$

and

$$\frac{\left(\lambda\left\{\left[I\left(u/T\right)+1\right]T+y\right\}\right)^k}{k!}\exp\left(-\lambda\left\{\left[I\left(u/T\right)+1\right]T+y\right\}\right). \tag{6}$$

Let us fix y and divide the service time into intervals of length T consisting of two parts, y and $T-y$. The probability (5) corresponds to the case when the mod T service time is less than y, the probability (6) when it is greater than y. The generating function of the number of entered customers has the form (provided the mod T interarrival time is y):

$$\mathbf{E}\{z^{\xi}|y\} = \sum_{k=0}^{\infty}\sum_{i=0}^{\infty}\left(\int_{iT}^{iT+y}\frac{[\lambda(iT+y)z]^k}{k!}e^{-\lambda(iT+y)}\mu e^{-\mu u}du + \right.$$
$$+\int_{iT+y}^{(i+1)T}\frac{[\lambda((i+1)T+y)z]^k}{k!}e^{-\lambda((i+1)T+y)}\mu e^{-\mu u}du =$$

(8)

$$= \frac{1}{1-e^{-[\lambda(1-z)+\mu]T}}\left(e^{-\lambda(1-z)y} - e^{-[\lambda(1-z)+\mu]y} + \right.$$
$$+e^{-\lambda(1-z)T}e^{-[\lambda(1-z)+\mu]y} - e^{-\lambda(1-z)y}e^{-[\lambda(1-z)+\mu]T},$$

(10)

where ξ is a random variable, the number of arriving customers for these periods. Multiplying this expression by $\frac{\lambda e^{-\lambda y}}{1-e^{-\lambda T}}$ and integrating from 0 to T, we obtain the generating function of transition probabilities

$$B(z) = \sum_{i=0}^{\infty} b_i z^i = \tag{10}$$
$$= \frac{1}{(1-e^{-\lambda T})(1-e^{-[\lambda(1-z)+\mu]T})}\left[\frac{1}{2-z}\left(1-e^{-\lambda(2-z)T}\right)\left(1-e^{-[\lambda(1-z)+\mu]T}\right) - \right.$$
$$\left. -\frac{\lambda}{\lambda(2-z)+\mu}\left(1-e^{-[\lambda(2-z)+\mu]T}\right)\left(1-e^{-\lambda(1-z)T}\right)\right].$$

Our results are collected in the following

Theorem. Let us consider a queueing system with Poisson arrivals with parameter λ, the service time is exponentially distributed with parameter μ. The service of a customer can start at the moment of its arrival (if the system is free) or at a moment that differs from it by the multiple of a given cycle time T (if the server is busy or there is a queue). The service discipline FCFS is accepted. If the server is idle, there is no customer arrived earlier, the current customer is at the corresponding position, the service necessarily begins. Let us introduce a Markov chain whose states correspond to the number of customers at moments $t_k - 0$ (t_k is the starting moment of service of the k-th customer). Its matrix of transition probabilities has the form

$$\begin{bmatrix} a_0 & a_1 & a_2 & a_3 & \dots \\ a_0 & a_1 & a_2 & a_3 & \dots \\ 0 & b_0 & b_1 & b_2 & \dots \\ 0 & 0 & b_0 & b_1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{12}$$

whose elements are determined by the generating functions (4) and (10). The generating function of ergodic distribution has the form

$$P(z) = p_0 \frac{B(z)(\lambda z + \mu) - zA(z)(\lambda + \mu)}{\mu[B(z) - z]}, \tag{17}$$

where

$$p_0 = 1 - \frac{\lambda}{\lambda + \mu} \frac{1 - e^{-(\lambda+\mu)T}}{e^{-\lambda T}(1 - e^{-\mu T})}. \tag{18}$$

The condition of existence of ergodic distribution is

$$\frac{\lambda}{\mu} < \frac{e^{-\lambda T}(1 - e^{-\mu T})}{1 - e^{-\lambda T}}. \tag{19}$$

Proof. The functioning of system can be described by the embedded Markov chain with the matrix of transition probabilities (16). Denote the ergodic probabilities by p_i $(i = 0, 1, \ldots)$ and introduce the generating function $P(z) = \sum_{i=0}^{\infty} p_i z^i$. We have

$$p_j = p_0 a_j + p_1 a_j + \sum_{i=2}^{j+1} p_i b_{j-i+1}, \tag{20}$$

$$p_0 = p_0 a_0 + p_1 a_0. \tag{21}$$

From (20) and (21)

$$P(z) = \frac{p_0 [zA(z) - B(z)] + p_1 z [A(z) - B(z)]}{z - B(z)}.$$

This expression includes two unknown probabilities p_0 and p_1 from the desired distribution, by (21) p_1 can be expressed by p_0, and p_0 can be found from the condition $P(1) = 1$, i.e.

$$p_0 = \frac{1 - B'(1)}{1 + A'(1) - B'(1) + \frac{\lambda}{\mu}[A' - B'(1)]}.$$

The chain is irreducible, so $p_0 > 0$. We have

$$\begin{aligned} A'(1) &= \frac{\lambda}{\lambda + \mu} \left(1 + \frac{\lambda T}{1 - e^{-\mu T}}\right), \\ B'(1) &= 1 - \frac{\lambda T e^{-\lambda T}}{1 - e^{-\lambda T}} + \frac{\lambda}{\lambda + \mu} \lambda T \frac{1 - e^{-(\lambda+\mu)T}}{(1 - e^{-\lambda T})(1 - e^{-\mu T})}, \end{aligned}$$

this yields

$$\left(1 + \frac{\lambda}{\mu}\right) A'(1) - \frac{\lambda}{\mu} B'(1) = \frac{\lambda}{\lambda + \mu} \lambda T \frac{1 - e^{-(\lambda+\mu)T}}{(1 - e^{-\lambda T})(1 - e^{-\mu T})} > 0,$$

so the condition $1 - B'(1) > 0$ must be fulfilled. This leads to the inequality

$$\frac{\lambda T e^{-\lambda T}}{1 - e^{-\lambda T}} - \frac{\lambda}{\lambda + \mu} \lambda T \frac{1 - e^{-(\lambda+\mu)T}}{(1 - e^{-\lambda T})(1 - e^{-\mu T})} > 0,$$

i.e.

$$\frac{\lambda}{\lambda+\mu} < \frac{e^{-\lambda T}(1-e^{-\mu T})}{1-e^{-(\lambda+\mu)T}},$$

which is equivalent to (19).

Between two customers there are idle periods to achieve the position to begin the service, as T decreases their influence decreases, too, in the limit case the service process becomes continuous.

Corollary. The limit distribution for the system described in the previous theorem as $T \to 0$ is

$$P^*(z) = \frac{1-\rho}{1-\rho z} \qquad \left(\rho = \frac{\lambda}{\mu}\right),$$

i.e. it is geometrical distribution with parameter ρ.

Proof. We find the limits p_0, $A(z)$ and $B(z)$ as $T \to 0$, and denote the limiting values by p_0^*, $A^*(z)$ and $B^*(z)$, respectively. We obtain

$$p_0^* = \lim_{T\to 0} p_0 = \lim_{T\to 0}\left(1 - \frac{\lambda}{\lambda+\mu}\frac{1-e^{-(\lambda+\mu)T}}{e^{-\lambda T}(1-e^{-\mu T})}\right) = 1 - \frac{\lambda}{\mu} = 1-\rho,$$

$$\begin{aligned} A^*(z) = \lim_{T\to 0} A(z) &= \lim_{T\to 0}\left(\frac{\mu}{\lambda+\mu} + \frac{\lambda z}{\lambda+\mu}\frac{e^{-\lambda(1-z)T}(1-e^{-\mu T})}{1-e^{-[\lambda(1-z)+\mu]T}}\right) = \\ &= \frac{\mu}{\lambda(1-z)+\mu}, \end{aligned}$$

$$\begin{aligned} B^*(z) &= \lim_{T\to 0} B(z) = \frac{1}{(1-e^{-\lambda T})[1-e^{-[\lambda(1-z)+\mu]T}]} \times \\ &\times \left\{ \frac{1}{2-z}\left(1-e^{-\lambda(2-z)T}\right)\left(1-e^{-[\lambda(1-z)+\mu]T}\right) - \right. \\ &\left. -\frac{\lambda}{\lambda(2-z)+\mu}\left(1-e^{-[\lambda(2-z)+\mu]T}\right)\left(1-e^{-\lambda(1-z)T}\right) \right\} = \\ &= \frac{\mu}{\lambda(1-z)+\mu}, \end{aligned}$$

and, by using these values,

$$P^*(z) = (1-\rho)\frac{(\lambda z+\mu)\frac{\mu}{\lambda(1-z)+\mu} - z(\lambda+\mu)\frac{\mu}{\lambda(1-z)+\mu}}{\mu\left(\frac{\mu}{\lambda(1-z)+\mu} - z\right)} = \frac{1-\rho}{1-\rho z}.$$

This is the generating function of ergodic distribution for the M/M/1 system, we come to the well-known classical result.

3.1.2. 3.1.2 The discrete time case

We are going to consider the discrete time version of the above described cyclic-waiting system. Let us divide the cycle time T into n equal parts and assume that for a time slice T/n a new customer arrives with probability r (so there is no entry with probability $1-r$), and the service of actual customer (if for this time slice it takes place) is continued with probability q and terminated with probability $1-q$. From these assumptions follows both the interarrival and service times have geometrical distributions.

Theorem. Let us consider a discrete queueing system in which both the interarrival and service time distributions are geometrical, the service of a customer may be started upon arrival or (in case of busy server or waiting queue) at moments differing from it by the multiples of a cycle time T equal to n time units. Let us define an embedded Markov chain whose states correspond to the number of customers in the system at moments t_k-0, where t_k is the moment of beginning of service of the k-th one. The matrix of transition probabilities has the form

$$\begin{bmatrix} a_0 & a_1 & a_2 & a_3 & \dots \\ a_0 & a_1 & a_2 & a_3 & \dots \\ 0 & b_0 & b_1 & b_2 & \dots \\ 0 & 0 & b_0 & b_1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

its elements are determined by the generating functions

$$A(z)=\sum_{i=0}^{\infty} a_i z^i = \frac{(1-r)(1-q)}{1-q(1-r)} + z\frac{r(1-q)}{1-q(1-r)} + z\frac{rq(1-r+rz)^n(1-q^n)}{[1-q(1-r)][1-q^n(1-r+rz)^n]},$$

$$B(z)=\sum_{k=1}^{n} \frac{(1-r)^{k-1}r}{1-(1-r)^n}\frac{1}{1-q^n(1-r+rz)^n}\times$$

$$\times[(1-r+rz)^k-(1-r+rz)^k q^k+(1-r+rz)^{n+k}q^k-(1-r+rz)^{n+k}q^n]=$$

$$=\frac{1-(1-r)^n(1-r+rz)^n}{1-(1-r)(1-r+rz)}\frac{r(1-r+rz)}{1-(1-r)^n}+$$

$$+\frac{1-q^n(1-r)^n(1-r+rz)^n}{1-q(1-r)(1-r+rz)}\frac{rq(1-r+rz)[(1-r+rz)^n-1]}{[1-(1-r)^n][1-q^n(1-r+rz)^n]}.$$

The generating function of ergodic distribution $P(z)=\sum_{i=0}^{\infty} p_i z^i$ has the form

$$P(z)=p_0\frac{zA(z)-B(z)+\frac{rz}{(1-r)(1-q)}[A(z)-B(z)]}{z-B(z)},$$

where

$$p_0=\frac{(1-r)(1-q)}{1-q(1-r)}-\frac{rq[1-(1-r)^n]}{(1-r)^{n-1}(1-q^n)[1-q(1-r)]}.$$

The condition of existence of ergodic distribution is the fulfilment of inequality

$$\frac{rq}{1-q^n}\frac{1-q^n(1-r)^n}{1-q(1-r)}<(1-r)^n.$$

Proof. Similarly to the continuous time case for the description of the system we will use an embedded Markov chain whose states correspond to the number of customers at moments t_k-0, i.e. at moments just before starting the service of k-th one. We find the transition probabilities for this chain.

Similarly to the continuous time case we will consider two possibilities: at the beginning of service there is one customer in the system or there are at least two customers in the system.

The case of one customer. We begin the service of the customer and after a certain time the second one arrives. Let u be the service time and the second customer appears at v time after the beginning of service. The remaining service time is ℓ ($\ell = 0, 1, 2, \dots$) with probability

$$P\{u - v = \ell\} = \sum_{k=\ell+1}^{\infty} q^{k-1}(1-q)(1-r)^{k-\ell-1} r = \frac{r(1-q)q^{\ell}}{1-q(1-r)}.$$

We find the time from the entry of second customer till the beginning of its service. It is 0 if the customer arrives during the last time slice of the first customers service, n if $u - v$ belongs to the interval $[1, n]$, $2n$ if $u - v \in [n+1, 2n]$, and, generally, in if $u - v \in [(i-1)n+1, in]$. The corresponding probabilities are

$$\sum_{\ell=(i-1)n+1}^{in} \frac{r(1-q)q^{\ell}}{1-q(1-r)} = \frac{rq}{1-q(1-r)} \left(q^{(i-1)n} - q^{in}\right).$$

The generating function of number of customers arriving for a time slice is $1 - r + rz$, so the generating function of customers entering for the waiting time is

$$\sum_{i=1}^{\infty} \frac{rq(1-q^n)}{1-q(1-r)} q^{(i-1)n}(1-r+rz)^{in} = \frac{rq(1-r+rz)^n(1-q^n)}{[1-q(1-r)][1-q^n(1-r+rz)^n]}.$$

Taking into account that the first customer obligatorily arrives and the waiting time may be equal to zero for the generating function of entering customers we obtain

$$A(z) = \sum_{i=0}^{\infty} a_i z^i = \frac{(1-r)(1-q)}{1-q(1-r)} + z\frac{r(1-q)}{1-q(1-r)} + z\frac{rq(1-r+rz)^n(1-q^n)}{[1-q(1-r)][1-q^n(1-r+rz)^n]},$$

where $\frac{(1-r)(1-q)}{1-q(1-r)}$ is the probability of event for the service of first customer no further customers arrive.

The case of at least two customers. At the beginning of service of first customer the second customer is present, too. Let $x = u - \left[\frac{u-1}{n}\right] n$ ($[x]$ denotes the integer part of x), and let y be the mod T interarrival time $(1 \le y \le n)$. The time elapsed between the starting moments of two successive customers is

$$\left[\frac{u-1}{n}\right] n + y \quad \text{if} \quad x \le y \quad \text{and} \quad \left(\left[\frac{u-1}{n}\right] + 1\right) n + y \quad \text{if} \quad x > y.$$

Let us fix y and consider the cycle $[in+1, (i+1)n]$. If the service of first customer ends till y (including y), then the time till the beginning of service of second customer is $in + y$ and the probability of this event is

$$\sum_{j=in+1}^{in+y} q^{j-1}(1-q) = q^{in} - q^{in+y},$$

in case $x > y$ the time is $(i+1)n + y$ and the probability is

$$\sum_{j=in+y+1}^{(i+1)n} q^{j-1}(1-q) = q^{in+y} - q^{(i+1)n}.$$

i changes from 0 to ∞ (the summation is extended for all possible values of service time), for fixed y the generating functions of entering customers in the two cases will be

$$\sum_{i=0}^{\infty}[q^{in}-q^{in+y}](1-r+rz)^{in+y}=\frac{(1-r+rz)^y}{1-q^n(1-r+rz)^n}-\frac{(1-r+rz)^y q^y}{1-q^n(1-r+rz)^n},$$

$$\sum_{i=0}^{\infty}[q^{in+y}-q^{in+n}](1-r+rz)^{in+n+y}=\frac{q^y(1-r+rz)^{n+y}}{1-q^n(1-r+rz)^n}-\frac{q^n(1-r+rz)^{n+y}}{1-q^n(1-r+rz)^n}.$$

y has truncated geometrical distribution, it takes on the value k $(k=1,2,\ldots,n)$ with probability $\frac{(1-r)^{k-1}r}{1-(1-r)^n}$.

Consequently, the generating function of transition probabilities is

$$B(z)=\sum_{k=1}^{n}\frac{(1-r)^{k-1}r}{1-(1-r)^n}\frac{1}{1-q^n(1-r+rz)^n}$$

$$\times[(1-r+rz)^k-(1-r+rz)^k q^k+(1-r+rz)^{n+k}q^k-(1-r+rz)^{n+k}q^n]$$

$$=\frac{1-(1-r)^n(1-r+rz)^n}{1-(1-r)(1-r+rz)}\frac{r(1-r+rz)}{1-(1-r)^n}$$

$$+\frac{1-q^n(1-r)^n(1-r+rz)^n}{1-q(1-r)(1-r+rz)}\frac{rq(1-r+rz)[(1-r+rz)^n-1]}{[1-(1-r)^n][1-q^n(1-r+rz)^n]}.$$

We have seen that, as in the continuous case, the length of interval between two successive starting moments is determined by the service time of first customer and the interarrival time of first and second customers, so they are independent random variables. By using the memoryless property of geometrical distribution, we obtain the number of customers in the system at moments just before the beginning of services constitute a Markov chain.

The system is considered at moments just before starting the services of customers. Let us denote the ergodic distribution by p_i $(i=0,1,2,\ldots)$ and introduce the generating function by $P(z)=\sum_{i=0}^{\infty}p_i z^i$. For p_i we have the system of equations

$$\begin{aligned} p_0 &= p_0a_0+p_1a_0,\\ p_j &= p_0a_j+p_1a_j+\sum_{i=2}^{j+1}p_ib_{j-i+1}, \end{aligned}$$

from which

$$\sum_{j=0}^{\infty}p_jz^j=p_0A(z)+p_1A(z)+\sum_{j=0}^{\infty}\sum_{i=2}^{j+1}p_ib_{j-i+1}z^j=$$

$$=\frac{1}{z}P(z)B(z)-\frac{1}{z}p_0B(z)+p_0A(z)+p_1A(z)-p_1B(z),$$

or

$$P(z) = \frac{p_0[zA(z) - B(z)] + p_1 z[A(z) - B(z)]}{z - B(z)}.$$

Since

$$p_0 = p_0 a_0 + p_1 a_0,$$

we have

$$p_1 = \frac{1 - a_0}{a_0} p_0 = \frac{r}{(1-r)(1-q)} p_0.$$

We find p_0 from the condition $P(1) = 1$

$$p_0 = \frac{1 - B'(1)}{1 - B'(1) + A'(1) + \frac{r}{(1-r)(1-q)}[A'(1) - B'(1)]}.$$

The chain is irreducible, so $p_0 > 0$.

Using the values

$$\begin{aligned} A'(1) &= \frac{r}{1 - q(1-r)} + \frac{nr^2 q}{[1 - q(1-r)](1 - q^n)}, \\ B'(1) &= 1 - \frac{nr(1-r)^n}{1 - (1-r)^n} + \frac{nr^2 q[1 - q^n(1-r)^n]}{(1-q^n)[1 - (1-r)^n][1 - q(1-r)]}, \end{aligned}$$

we obtain

$$\left(1 + \frac{r}{(1-r)(1-q)}\right) A'(1) - \frac{r}{(1-r)(1-q)} B'(1)$$

$$= \frac{nr^2 q}{(1-q^n)[1 - q(1-r)]} + \frac{nr^2(1-r)^n}{(1-r)[1 - (1-r)^n][1 - q(1-r)]} > 0,$$

so the condition $1 - B'(1) > 0$ must be fulfilled. This leads to the expression

$$\frac{nr(1-r)^n}{1 - (1-r)^n} - \frac{nr^2 q[1 - q^n(1-r)^n]}{(1-q^n)[1 - (1-r)^n][1 - q(1-r)]} > 0.$$

From it we obtain the stability condition

$$\frac{rq}{1 - q^n} \frac{1 - q^n(1-r)^n}{1 - q(1-r)} < (1-r)^n.$$

The left side of this inequality is equal to

$$rq(1 + q^n + q^{2n} + \ldots)[1 + q(1-r) + q^2(1-r)^2 + \ldots + q^{n-1}(1-r)^{n-1}],$$

so it is continuous monotone decreasing function taking on values from $(\infty, 0)$ as $q \to 0$; for fixed n and arrival probability r one can obtain the possible values of q (it is necessary to decrease the value of q till the inequality becomes fulfilled.

3.2. 3.2 Waiting time

3.2.1. 3.2.1 The continuous time case

We consider the queueing system described in the previous part and use Kobas results to find the waiting time distribution.

Let t_n denote the moment of arrival of the n -th customer; its service will begin at the moment $t_n + T \cdot X_n$, where X_n is a nonnegative integer. Let $\xi_n = t_{n+1} - t_n$, and η_n be the service time of n -th customer. Furthermore, let $X_n = i$, if

$$(k-1)T < iT + Y_n - Z_n \leq kT \qquad (k \geq 1),$$

then $X_{n+1} = k$, and if $iT + \eta_n - \xi_n \leq 0$, then $X_{n+1} = 0$. Hence, X_n is a homogeneous Markov chain with transition probabilities p_{ik} , where

$$p_{ik} = P\{(k-i-1)T < \eta_n - \xi_n \leq (k-i)T\}$$

if $k \geq 1$, and

$$p_{i0} = P\{\eta_n - \xi_n \leq -iT\}.$$

Introduce the notations

$$f_j \;=\; P\{(j-1)T < \eta_n - \xi_n \leq jT\}, \tag{22}$$

$$p_{ik} \;=\; f_{k-i} \quad \text{if} \quad k \geq 1, \quad p_{i0} = \sum_{j=-\infty}^{-i} f_j = \hat{f}_i. \tag{23}$$

The ergodic distribution of this chain satisfies the system of equations

$$\begin{aligned} p_j &= \sum_{i=0}^{\infty} p_i p_{ij} \qquad (j \geq 0), \\ \sum_{j=0}^{\infty} p_j &= 1. \end{aligned}$$

Theorem. Let us consider the system described earlier and introduce a Markov chain whose states correspond to the waiting time (in the sense that the waiting time is the number of actual state multiplied by T) at the arrival time of customers. The matrix of transition probabilities for this chain is

$$\begin{bmatrix} \sum\limits_{j=-\infty}^{0} f_j & f_1 & f_2 & f_3 & f_4 & \cdots \\ \sum\limits_{j=-\infty}^{-1} f_j & f_0 & f_1 & f_2 & f_3 & \cdots \\ \sum\limits_{j=-\infty}^{-2} f_j & f_{-1} & f_0 & f_1 & f_2 & \cdots \\ \sum\limits_{j=-\infty}^{-3} f_j & f_{-2} & f_{-1} & f_0 & f_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{24}$$

its elements are defined by (22) and (23). The generating function of the ergodic distribution is

$$P(z) = \left[1 - \frac{\lambda}{\mu}\frac{1-e^{-\lambda T}}{e^{-\lambda T}(1-e^{-\mu T})}\right] \times \times \frac{\frac{\mu}{\lambda+\mu} - \frac{\mu(1-e^{-\lambda T})}{\lambda+\mu}\frac{z}{z-e^{-\lambda T}}}{1 - \frac{\lambda(1-e^{-\mu T})}{\lambda+\mu}\frac{z}{1-ze^{-\mu T}} - \frac{\mu(1-e^{-\lambda T})}{\lambda+\mu}\frac{z}{z-e^{-\lambda T}}}, \quad (29)$$

the condition of existence of ergodic distribution is

$$\frac{\lambda}{\mu} < \frac{e^{-\lambda T}(1-e^{-\mu T})}{1-e^{-\lambda T}}. \quad (30)$$

Proof. For the system we have

$$P\{\eta < x\} = 1 - e^{-\lambda x}, \qquad P\{\xi < x\} = 1 - e^{-\mu x}.$$

Let us find the distribution of $\eta - \xi$. Let $x \geq 0$, then $\eta < \xi + x$ and the probability of this event is

$$\int_0^\infty [1 - e^{-\mu(y+x)}]\lambda e^{-\lambda y} dy = 1 - \frac{\lambda}{\lambda+\mu}e^{-\mu x}.$$

Let $x < 0$, then $\xi > \eta + |x|$, from which we obtain the probability

$$\int_0^\infty e^{-\lambda(y+|x|)}\mu e^{-\mu y} dy = \frac{\mu}{\lambda+\mu}e^{-\lambda|x|} = \frac{\mu}{\lambda+\mu}e^{\lambda x}.$$

So, the distribution function of $\eta - \xi$ is

$$F(x) = \begin{cases} \frac{\mu}{\lambda+\mu}e^{\lambda x} & \text{if} x \leq 0, \\ 1 - \frac{\lambda}{\lambda+\mu}e^{-\mu x} & \text{if } x > 0. \end{cases}$$

The transition probabilities of the Markov chain are, if $j > 0$

$$f_j = 1 - \frac{\lambda}{\lambda+\mu}e^{-\mu jT} - 1 + \frac{\lambda}{\lambda+\mu}e^{-\mu(j-1)T} = \frac{\lambda}{\lambda+\mu}(1-e^{-\mu T})e^{-\mu(j-1)T},$$

for the negative values $(j \geq 0)$

$$f_{-j} = \frac{\mu}{\lambda+\mu}e^{-\lambda jT} - \frac{\mu}{\lambda+\mu}e^{-\lambda(j+1)T} = \frac{\mu}{\lambda+\mu}(1-e^{-\lambda T})e^{-\lambda jT},$$

$$p_{i0} = \hat{f}_i = \sum_{j=-\infty}^{-i} f_j = \sum_{j=i}^{\infty} \frac{\mu}{\lambda+\mu}(1-e^{-\lambda T})e^{-\lambda jT} = \frac{\mu}{\lambda+\mu}e^{-\lambda iT}.$$

Using the matrix of transition probabilities, we obtain the system of equations

$$\begin{aligned} p_0 &= p_0\hat{f}_0 + p_1\hat{f}_1 + p_2\hat{f}_2 + p_3\hat{f}_3 + \dots \\ p_1 &= p_0 f_1 + p_1 f_0 + p_2 f_{-1} + p_3 f_{-2} + \dots \\ p_2 &= p_0 f_2 + p_1 f_1 + p_2 f_0 + p_3 f_{-1} + \dots \\ &\vdots \end{aligned}$$

Multiplying the j -th equation by z^j , summing up from zero to infinity, for the generating function $\sum_{j=0}^{\infty} p_j z^j$ we have

$$P(z) = P(z)F_+(z) + \sum_{j=1}^{\infty} p_j z^j \sum_{i=0}^{j-1} f_{-i} z^{-i} + \sum_{j=0}^{\infty} p_j \hat{f}_j. \quad (31)$$

For our system

$$\begin{aligned} F_+(z) &= \sum_{i=1}^{\infty} f_i z^i = \frac{\lambda z}{\lambda+\mu}(1-e^{-\mu T}) \sum_{i=1}^{\infty} e^{-\mu(i-1)T} z^{i-1} = \\ &= \frac{\lambda(1-e^{-\mu T})}{\lambda+\mu} \frac{z}{1-ze^{-\mu T}}, \\ \sum_{i=0}^{j-1} f_{-i} z^{-i} &= \frac{\mu(1-e^{-\lambda T})}{\lambda+\mu} \sum_{i=0}^{j-1} e^{-\lambda i T} z^{-i} = \frac{\mu(1-e^{-\lambda T})}{\lambda+\mu} \frac{1-\left(\frac{e^{-\lambda T}}{z}\right)^j}{1-\frac{e^{-\lambda T}}{z}}, \\ \sum_{i=0}^{\infty} p_i \hat{f}_i &= \sum_{i=0}^{\infty} p_i \frac{\mu}{\lambda+\mu} e^{-\lambda i T} = \frac{\mu}{\lambda+\mu} P\left(e^{-\lambda T}\right). \end{aligned}$$

Substituting these expressions into (31) yields

$$\begin{aligned} P(z) &= P(z)F_+(z) + \sum_{j=1}^{\infty} p_j z^j \frac{\mu(1-e^{-\lambda T})}{\lambda+\mu} \frac{1-\left(\frac{e^{-\lambda T}}{z}\right)^j}{1-\frac{e^{-\lambda T}}{z}} + \frac{\mu}{\lambda+\mu} P\left(e^{-\lambda T}\right) = \\ &= P(z)F_+(z) + \frac{\mu(1-e^{-\lambda T})}{\lambda+\mu} \frac{z}{z-e^{-\lambda T}} \left[P(z) - P\left(e^{-\lambda T}\right)\right] + \frac{\mu}{\lambda+\mu} P\left(e^{-\lambda T}\right), \end{aligned}$$

or

$$P(z)\left[1 - F_+(z) - \frac{\mu(1-e^{-\lambda T})}{\lambda+\mu} \frac{z}{z-e^{-\lambda T}}\right] = P\left(e^{-\lambda T}\right)\left[\frac{\mu}{\lambda+\mu} - \frac{\mu(1-e^{-\lambda T})}{\lambda+\mu} \frac{z}{z-e^{-\lambda T}}\right]$$

The value of $P(e^{-\lambda T})$ can be found from the fact $P(1) = 1$,

$$P\left(e^{-\lambda T}\right) = 1 - \frac{\lambda}{\mu} \frac{1-e^{-\lambda T}}{e^{-\lambda T}(1-e^{-\mu T})}.$$

For the generating function of waiting time we obtain the above expression, whence the probability of zero waiting time is

$$p_0 = \left[1 - \frac{\lambda}{\mu} \frac{1-e^{-\lambda T}}{e^{-\lambda T}(1-e^{-\mu T})}\right] \frac{\mu}{\lambda+\mu}.$$

Because of ergodicity $p_0 > 0$ must hold, so the inequality

$$\frac{\lambda}{\mu}\frac{1-e^{-\lambda T}}{e^{-\lambda T}(1-e^{-\mu T})} < 1$$

must be fulfilled. It leads to the condition (30), and coincides with the stability condition for the number of customers.

We find the mean value of waiting time. The generating function of waiting time (measured in cycles) is

$$\begin{aligned} P(z) &= \left[1-\frac{\lambda}{\mu}\frac{1-e^{-\lambda T}}{e^{-\lambda T}(1-e^{-\mu T})}\right]\times \\ &\times\frac{\frac{\mu}{\lambda+\mu}-\frac{\mu(1-e^{-\lambda T})}{\lambda+\mu}\frac{z}{z-e^{-\lambda T}}}{1-\frac{\lambda(1-e^{-\mu T})}{\lambda+\mu}\frac{z}{1-ze^{-\mu T}}-\frac{\mu(1-e^{-\lambda T})}{\lambda+\mu}\frac{z}{z-e^{-\lambda T}}}. \end{aligned}$$

Introducing the notations

$$\begin{aligned} A(z) &= \frac{\mu}{\lambda+\mu}-\frac{\mu(1-e^{-\lambda T})}{\lambda+\mu}\frac{z}{z-e^{-\lambda T}}, \\ B(z) &= 1-\frac{\lambda(1-e^{-\mu T})}{\lambda+\mu}\frac{z}{1-ze^{-\mu T}}-\frac{\mu(1-e^{-\lambda T})}{\lambda+\mu}\frac{z}{z-e^{-\lambda T}}, \end{aligned}$$

the mean value of number of cycles is

$$\lim_{z\to 1}\left[1-\frac{\lambda}{\mu}\frac{1-e^{-\lambda T}}{e^{-\lambda T}(1-e^{-\mu T})}\right]\frac{A'B-AB'}{2B^2}.$$

By using twice lHospitals rule and taking into account

$$\lim_{z\to 1}\frac{A'B-AB'}{2B^2}=\lim_{z\to 1}\frac{A''B'-A'B''}{2B'^2},$$

$$\begin{aligned} A'(1) &= \frac{\mu e^{-\lambda T}}{(\lambda+\mu)(1-e^{-\lambda T})}, \\ A''(1) &= -\frac{2\mu e^{-\lambda T}}{(\lambda+\mu)(1-e^{-\lambda T})^2}, \\ B'(1) &= -\frac{\lambda}{(\lambda+\mu)(1-e^{-\mu T})}+\frac{\mu e^{-\lambda T}}{(\lambda+\mu)(1-e^{-\lambda T})}, \\ B''(1) &= -\frac{2\lambda e^{-\mu T}}{(\lambda+\mu)(1-e^{-\mu T})^2}-\frac{2\mu e^{-\lambda T}}{(\lambda+\mu)(1-e^{-\lambda T})^2}, \end{aligned}$$

we finally obtain

$$P'(1)=\frac{\lambda[1-e^{-(\lambda+\mu)T}]}{(1-e^{-\mu T})[\mu e^{-\lambda T}(1-e^{-\mu T})-\lambda(1-e^{-\lambda T})]}.$$

3.2.2. 3.2.2 The discrete time case

We determine the distribution of $\eta - \xi$ if both of them are geometrically distributed. The probability that $\eta - \xi = j > 0$ is equal to

$$\sum_{i=1}^{\infty}(1-r)^{i-1} r q^{i-1+j}(1-q) = \frac{r(1-q)(1-q)^j}{1-q(1-r)} \qquad (j=1,2\ldots)$$

if $\eta - \xi > 0$, and

$$\sum_{i=1}^{\infty} q^{i-1}(1-q)(1-r)^{i-1+j} r = \frac{r(1-q)(1-r)^j}{1-q(1-r)} \qquad (j=0,1,2,\ldots)$$

if $\eta - \xi = -j \leq 0$. The transition probabilities are given by the formulas

$$f_j = \sum_{k=(j-1)n+1}^{jn} \frac{r(1-q)(1-q)^k}{1-q(1-r)} = \frac{rq(1-q^n)}{1-q(1-r)} q^{(j-1)n}$$

in case of positive jumps, and

$$f_{-j} = \sum_{k=jn}^{(j+1)n-1} \frac{r(1-q)(1-r)^k}{1-q(1-r)} = \frac{(1-q)[1-(1-r)^n]}{1-q(1-r)}(1-r)^{jn}$$

for the case of nonpositive jumps. Furthermore, we have

$$p_{j0} = \sum_{k=-\infty}^{-j} f_k = \sum_{k=j}^{\infty} \frac{(1-q)[1-(1-r)^n]}{1-q(1-r)}(1-r)^{kn} = \frac{(1-q)(1-r)^{jn}}{1-q(1-r)} = \hat{f}_j.$$

By using these transition probabilities for the equilibrium distribution we have the system of linear equations

$$\begin{aligned}
p_0 &= p_0\hat{f}_0 + p_1\hat{f}_1 + p_2\hat{f}_2 + p_3\hat{f}_3 + \ldots \\
p_1 &= p_0 f_1 + p_1 f_0 + p_2 f_{-1} + p_3 f_{-2} + \ldots \\
p_2 &= p_0 f_2 + p_1 f_1 + p_2 f_0 + p_3 f_{-1} + \ldots \\
&\vdots
\end{aligned}$$

Multiplying the j -th equation by z^j , summing up from zero to infinity, for the generating function $P(z) = \sum_{j=0}^{\infty} p_j z^j$ we have

$$P(z) = P(z)F_+(z) + \sum_{j=1}^{\infty} p_j z^j \sum_{i=0}^{j-1} f_{-i} z^{-i} + \sum_{j=0}^{\infty} p_j \hat{f}_j,$$

where

$$F_+(z) = \sum_{i=1}^{\infty} f_i z^i.$$

By using the transition probabilities f_j we have

$$F_+(z) = \sum_{j=1}^{\infty} f_j z^j = \sum_{j=1}^{\infty} \frac{rq(1-q^n)}{1-q(1-r)} q^{(j-1)n} z^j =$$

$$= \frac{rq(1-q^n)}{1-q(1-r)} \frac{z}{1-q^n z},$$

$$\begin{aligned} \sum_{i=0}^{j-1} f_{-i} z^{-i} &= \sum_{i=0}^{j-1} \frac{(1-q)[1-(1-r)^n]}{1-q(1-r)} \left(\frac{(1-r)^n}{z}\right)^i = \\ &= \frac{(1-q)[1-(1-r)^n]}{1-q(1-r)} \frac{1-\left(\frac{(1-r)^n}{z}\right)^j}{1-\frac{(1-r)^n}{z}} = \\ &= \frac{(1-q)[1-(1-r)^n]}{1-q(1-r)} \frac{z}{z-(1-r)^n} \left[1-\left(\frac{(1-r)^n}{z}\right)^j\right], \end{aligned}$$

$$\begin{aligned} \sum_{i=1}^{\infty} p_i \hat{f}_i &= \sum_{i=0}^{\infty} p_i \frac{1-q}{1-q(1-r)} (1-r)^{in} = \\ &= \frac{1-q}{1-q(1-r)} \sum_{i=0}^{\infty} p_i (1-r)^{in} = \frac{1-q}{1-q(1-r)} P((1-r)^n), \end{aligned}$$

$$\sum_{j=1}^{\infty} p_j z^j \sum_{i=0}^{j-1} f_{-i} z^{-i} = \sum_{j=1}^{\infty} p_j z^j \frac{(1-q)[1-(1-r)^n]}{1-q(1-r)} \frac{z}{z-(1-r)^n} \left[1-\left(\frac{(1-r)^n}{z}\right)^j\right] =$$

$$\begin{aligned} &= \frac{(1-q)[1-(1-r)^n]}{1-q(1-r)} \frac{z}{z-(1-r)^n} \sum_{j=1}^{\infty} p_j z^j \left[1-\left(\frac{(1-r)^n}{z}\right)^j\right] = \\ &= \frac{(1-q)[1-(1-r)^n]}{1-q(1-r)} \frac{z}{z-(1-r)^n} [P(z)-P((1-r)^n)]. \end{aligned}$$

So, the expression for the generating function may be written in the form

$$P(z)\left[1-\frac{rq(1-q^n)}{1-q(1-r)} \frac{z}{1-q^n z} - -\frac{(1-q)[1-(1-r)^n]}{1-q(1-r)} \frac{z}{z-(1-r)^n}\right] =$$

$$= P((1-r)^n)\left[\frac{1-q}{1-q(1-r)} - \frac{(1-q)[1-(1-r)^n]}{1-q(1-r)} \frac{z}{z-(1-r)^n}\right].$$

This expression contains the unknown value $P((1-r)^n)$, it can be found from the condition $P(1)=1$. It is equal to

$$P((1-r)^n) = 1-\frac{rq[1-(1-r)^n]}{(1-q)(1-q^n)(1-r)^n},$$

so, finally, the generating function takes on the form

$$P(z) = \left[1-\frac{rq[1-(1-r)^n]}{(1-q)(1-q^n)(1-r)^n}\right] \times$$

$$\times \frac{\frac{1-q}{1-q(1-r)} - \frac{(1-q)[1-(1-r)^n]}{1-q(1-r)} \frac{z}{z-(1-r)^n}}{1 - \frac{rq(1-q^n)}{1-q(1-r)} \frac{z}{1-q^n z} - \frac{(1-q)[1-(1-r)^n]}{(1-q)(1-q^n)(1-r)^n} \frac{z}{z-(1-r)^n}}.$$

The chain is irreducible and aperiodic, in order to get the stability condition we find p_0 from the generating function

$$p_0 = \left[1 - \frac{rq[1-(1-r)^n]}{(1-q)(1-q^n)(1-r)^n}\right] \frac{1-q}{1-q(1-r)}.$$

It is positive if

$$\frac{rq[1-(1-r)^n]}{(1-q)(1-q^n)(1-r)^n} < 1,$$

i.e. it is equivalent to the stability condition in the case of the number of customers.

4. 4 Markov chains

Let $E_1, E_2, \ldots, E_i, \ldots$ be a complete group of mutually exclusive events, ξ_n $(n = 0, 1, 2, \ldots)$ a sequence of random variables. $\xi_n = i$ in the n -th trial the event E_i is realized. In case of independent trials

$$P\{\xi_n = j \mid \xi_0 = i_0, \xi_1 = i_1, \xi_2 = i_2, \ldots, \xi_{n-1} = i_{n-1}\} = P\{\xi_n = j\}.$$

If for all n and all possible values of variables

$$P\{\xi_n = j \mid \xi_0 = i_0, \xi_1 = i_1, \ldots, \xi_{n-1} = i_{n-1}\} = P\{\xi_n = j \mid \xi_{n-1} = i_{n-1}\},$$

then the random variables constitute a Markov chain.

The distribution $P\{\xi_0 = i\} = P_i(0)$ is called the initial distribution of the random variable ξ_0, the conditional probabilities $P\{\xi_n = j \mid \xi_{n-1} = i\}$ are called transition probabilities.

If we know the initial distribution and the transition probabilities, they determine the distribution of ξ_n on a unique way.

If the transition probabilities

$$P\{\xi_n = j \mid \xi_{n-1} = i\} = p_{ij}.$$

do not depend on n then the Markov chain is homogeneous. The transition probabilities may be written in the form of a matrix

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ p_{31} & p_{32} & p_{33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

where obviously

$$p_{i1} + p_{i2} + p_{i3} + \ldots = 1.$$

Let $P_{ij}(n)$ be the probability of event that the system from state i comes to state j for n steps. By the formula of total probability

$$P_{ij}(n) = \sum_{r=1}^{n} P_{ir}(m) P_{rj}(n-m).$$

Let π_n denote the matrix the matrix of n -step transition probabilities

$$\pi_n = \begin{bmatrix} P_{11}(n) & P_{12}(n) & P_{13}(n) & \dots \\ P_{21}(n) & P_{22}(n) & P_{23}(n) & \dots \\ P_{31}(n) & P_{32}(n) & P_{33}(n) & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

By using the relation for the n -step transition probability we have

$$\pi_n = \pi_m \cdot \pi_{n-m} \qquad (0 < m < n).$$

In the case $n = 2$

$$\pi_2 = \pi_1 \cdot \pi_1 = \pi_1^2,$$

in the case $n = 3$

$$\pi_3 = \pi_1 \cdot \pi_2 = \pi_2 \cdot \pi_1 = \pi_1^3,$$

and, generally,

$$\pi_n = \pi_1^n.$$

The classification of states of the Markov chain. The state E_k can be reached from the state E_j if for a certain $n > 0$ the inequality $p_{jk}^{(n)} > 0$ holds. The Markov chain is irreducible if any state can be reached from any state for a certain number of steps.

Let us consider a fixed state E_j. Let $f_j^{(n)}$ be the probability of first passage from the state E_j to the state E_j in the n -th step. Then we come from the state E_j to the same state with probability

$$p_{jj}^{(n)} = \sum_{m=1}^{n} f_j^{(m)} p_{jj}^{(n-m)}.$$

The probability to return to the state j at all is

$$f_j = \sum_{n=1}^{\infty} f_j^{(n)}.$$

If $f_j = 1$, the system obligatorily returns to the state j. The average return time is

$$\mu_j = \sum_{n=1}^{\infty} n f_j^{(n)}.$$

The state E_j is called recurrent if $f_j = 1$. The state E_j is transient if $f_j < 1$.

The recurrent state E_j is called zero state if the mean value of return time is infinite.

The state E_j is periodic with period $t > 1$ if the return is possible only at moments $t, 2t, 3t, \ldots$.

The recurrent state E_j is ergodic if it is not zero state and it is aperiodic.

The states of an irreducible Markov chain always belong to the same class: they are either transient, or recurrent zero or ergodic.

Theorem. If the states of an irreducible Markov chain are aperiodic, non-transient and non-zero states then independently of the initial distribution $P_j(0)$ there exist the limiting probabilities

$$\lim_{n\to\infty} P_j(n) = P_j$$

and

$$\sum_j P_j = 1.$$

The distribution $\{P_j\}$ is uniquely determined by the system of linear equations

$$P_j = \sum_i P_i p_{ij}.$$

5. References