




Article

Computation of Transient and Steady-State Characteristics of Queueing Systems with Different Types of Customer

Alexander Zeifman ^{1,2} , Yacov Satin ¹ , Ilia Usov ¹ and Janos Sztrik ^{3,*} 

¹ Department of Applied Mathematics, Vologda State University, 160000 Vologda, Russia; a_zeifman@mail.ru (A.Z.); yacovi@mail.ru (Y.S.); iusov35@yandex.ru (I.U.)

² Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 119133 Moscow, Russia

³ Faculty of Informatics, University of Debrecen, 4032 Debrecen, Hungary

* Correspondence: sztrik.janos@inf.unideb.hu

Abstract: This paper deals with queueing models, in which the number of customers is described by a (inhomogeneous, in general) birth–death process. Depending on the choice of the type of intensities for the arrival and service of customers, the system can either have impatience (in which, as the queue length increases, the intensities of arrival decrease and the intensities of service increases) or attraction (in which, on the contrary, as the queue length increases, the intensities of the arrival of customers increase and service intensities decrease). In this article, various types of such models are considered, and their transient and limiting characteristics are computed. Furthermore, the rate of convergence and related bounds are also dealt with. Several numerical examples illustrate the proposed procedures.

Keywords: inhomogeneous continuous-time Markov chains; birth–death process; ergodicity; logarithmic norm

1. Introduction

This paper aims to show that, using (inhomogeneous) birth–death processes (BDPs) with state-dependent intensities, one can describe and investigate simple queueing theory models with different types of customer and server behavior.

In both the case of ergodicity and the case of null ergodicity, it is possible to obtain explicit estimates of the rate of convergence. In the ergodic case, these estimates also allow us to study the boundaries of perturbation and approximations using truncations and to compute the main limiting characteristics of the corresponding model. Thus, by selecting the model parameters, it is possible to control the queueing system.

We consider a general queueing system of type $M_n(t)/M_n(t)/1$, in which the intensities of the arrival and service of customers characterize the corresponding behavior (of customers and servers) in the queue.

The earliest models relate to queues where potential customers are discouraged by queue length; see [1–4]. This is reflected in the corresponding choice of monotonically decreasing intensities for the arrival of customers. Models with impatient customers began to be actively studied, in which the property of impatience was most often described by an increase in service intensities; see, for example, [5–11].



Received: 13 May 2025
Revised: 2 June 2025
Accepted: 10 June 2025
Published: 14 June 2025

Citation: Zeifman, A.; Satin, Y.; Usov, I.; Sztrik, J. Computation of Transient and Steady-State Characteristics of Queueing Systems with Different Types of Customer. *Computation* **2025**, *13*, 150. <https://doi.org/10.3390/computation13060150>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Further, queues with encouragement or with a server that works less effectively as the queue length increases are described by processes with increasing intensity of customer arrivals as the queue grows, or decreasing intensity of service; see [12,13]. As a result, the influence of the nature of the queue on the corresponding BDP can be taken into account due to the corresponding behavior of the intensities of the arrival/service of requests.

As one can see, most studies by other authors are devoted to constructing a particular solution at the initial stage of the system’s existence, usually for stationary systems. Often, they simply obtain a solution using purely numerical methods—uniformity, modeling, and Runge–Kutta methods—which, in general, do not provide useful information about the system; therefore, unlike the listed studies, the goal of this article is to construct a LIMIT solution, i.e., a stable solution that is practically independent of the initial state. Without knowing the convergence rate, it is not known when the limit solution has been reached. Comparison at the end of the period or (at constant intensities) of some interval is not enough to know whether the limit solution has been reached. In particular (for 1-- periodicity), to achieve the limit regime, it is necessary to cut the system into the interval $[0, n]$ until $\frac{1-\alpha}{\alpha} \|\mathbf{p}(n-1) - \mathbf{p}(n)\|$ is not less than the required accuracy. Here, α is the convergence rate parameter. Another approach is to calculate the value of t_* at which the limit solution is found. A table with a series of examples comparing these approaches is presented below. Note that, as the graphs below show, the actual value of t_* is usually less than the expected value. It is worth noting that a comparison of numerical methods can be found in studies from the 1970s. Regarding relatively modern ones, we can use [14]. The Runge–Kutta method is distinguished by its guaranteed accuracy and is therefore often chosen as a reference in numerical experiments. In particular, in this work, we completed our calculations using it.

The main role in this work is played by the examples of different types of BDPs considered, describing the corresponding models; for each type of system, the derivation of the estimates of the rate of convergence and the construction of the main limiting characteristics are shown. Perturbation bounds and approximation errors using truncations can then be formally obtained by the methods described in [15].

2. Model Description and Main Results

Let $M_n/M_n/1$ be a queueing system with arrival and service rates λ_n and μ_n , respectively, and let $X(t)$ be the corresponding queue-length process. Then, $X(t)$ is a BDP with infinite state space $\{0, 1, 2, \dots\}$, intensity matrix Q , and transposed intensity matrix $A = Q^T$. Then, matrix $A = (a_{ij})_{i,j=0}^\infty$ has the following form:

$$A = \begin{pmatrix} -\lambda_0 & \mu_1 & 0 & \dots \\ \lambda_0 & -(\lambda_1 + \mu_1) & \mu_2 & \dots \\ 0 & \lambda_1 & -(\lambda_2 + \mu_2) & \dots \\ 0 & 0 & \lambda_2 & \dots \\ 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{1}$$

We will assume that the arrival and service intensities satisfy the following conditions: $\lambda_i > 0 (i \geq 0)$, $\mu_i > 0 (i > 0)$, $\mu_0 = 0$, and $\sup_i (\lambda_i + \mu_i) < \infty$.

Denote the l_1 -norm by $\|\cdot\|$, with $\|\mathbf{x}\| = \sum |x_i|$ and $\|B\| = \sup_j \sum_i |b_{ij}|$ for matrix $B = (b_{ij})$.

We consider the corresponding forward Kolmogorov system:

$$\frac{d\mathbf{p}}{dt} = A\mathbf{p}(t), \tag{2}$$

as a differential equation in the space of sequences l_1 with bounded operator function A . Here, $\mathbf{p}(t) = (p_0(t), p_1(t), \dots)^T$ is the vector of state probabilities for BDP $X(t)$.

Moreover, if the queue is nonstationary, then the corresponding BDP is inhomogeneous and $\lambda_i(t)$ and $\mu_i(t)$ are locally integrable on $[0, \infty)$ nonnegative functions such that $\sup_i(\lambda_i(t) + \mu_i(t)) \leq L < \infty$ for almost all $t \geq 0$. In this situation, the operator function $A(t)$ is also bounded and locally integrable on $[0, \infty)$.

Recall that

A Markov chain $X(t)$ is called null ergodic if $p_n(t) \rightarrow 0$ as $t \rightarrow \infty$ for any n and any initial condition $\mathbf{p}(0)$;

A Markov chain $X(t)$ is called ergodic if it has a stationary distribution π such that $\|\mathbf{p}(t) - \pi\| \rightarrow 0$ as $t \rightarrow \infty$ for any initial condition $\mathbf{p}(0)$;

An inhomogeneous Markov chain $X(t)$ is called weakly ergodic if $\|\mathbf{p}^*(t) - \mathbf{p}^{**}(t)\| \rightarrow 0$ as $t \rightarrow \infty$ for any initial conditions $\mathbf{p}^*(0)$ and $\mathbf{p}^{**}(0)$.

Also recall that the logarithmic norm γ of the operator function $B(t)$ on l_1 gives us bounds for the rate of convergence for solutions of the corresponding linear differential equation and can be written as $\gamma(B(t)) = \sup_i (b_{ii}(t) + \sum_{j \neq i} |b_{ji}(t)|)$; see all details in [15].

Consider an auxiliary “weight” sequence $\{d_k\}$ such that $\inf_k(d_k) = d > 0$ and the corresponding upper triangular matrix D :

$$D = \begin{pmatrix} d_1 & d_1 & d_1 & \cdots \\ 0 & d_2 & d_2 & \cdots \\ 0 & 0 & d_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{3}$$

We write $\|\mathbf{z}\|_{1D} = \|D\mathbf{z}\|_1$ for a vector $\mathbf{z} = (z_1, z_2, \dots)$.

Set

$$\alpha_i = \lambda_i + \mu_{i+1} - \frac{d_{i+1}}{d_i} \lambda_{i+1} - \frac{d_{i-1}}{d_i} \mu_i, \tag{4}$$

and

$$\beta_* = \inf_{i \geq 0} \alpha_i.$$

Theorem 1. Let $X(t)$ be a homogeneous BDP with positive intensities λ_k and μ_k . Let there exist limits $\lambda^* = \lim_{k \rightarrow \infty} \lambda_k$ and $\lim_{k \rightarrow \infty} \mu_k = \mu^*$. Let $\lambda^* < \mu^*$. Then, the BDP is ergodic and there exists a sequence of positive numbers $\{d_i\}$ such that $\inf d_k = d > 0$, $\beta_* = \inf_i \alpha_i > 0$, and $\|p(t) - \pi\|_{1D} \leq e^{-\beta_* t} \|p(0) - \pi\|_{1D}$.

Proof. First, from our general scheme, we will carry out the following actions. From equality $\sum_{i \geq 0} p_i = 1$, we obtain $p_0 = 1 - \sum_{i \geq 1} p_i$. Hence, we have from (2) the following equation:

$$\frac{d\mathbf{z}}{dt} = B\mathbf{z} + \mathbf{f}, \tag{5}$$

where $\mathbf{z} = (p_1, p_2, \dots)^T$, $\mathbf{f} = (\lambda_0, 0, 0, \dots)^T$ and

$$B = \begin{pmatrix} -(\lambda_0 + \lambda_1 + \mu_1) & \mu_2 - \lambda_0 & -\lambda_0 & -\lambda_0 \cdots \\ 0 & \lambda_1 & -(\lambda_2 + \mu_2) & \cdots \\ 0 & 0 & \lambda_2 & \cdots \\ 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

For bounding the rate of convergence, we consider the respective ‘homogeneous’ equation $\frac{dx}{dt} = Bx$ in the corresponding space l_{1D} .

Then,

$$\gamma(B)_{1D} = \gamma(DBD^{-1})_1 = -\inf_{i \geq 0} \alpha_i = -\beta_*. \tag{6}$$

Now, we try to find a sequence $\{d_i\}$ such that all α_i are the same and positive. The following reasoning is close to that carried out in [16]. Let $\delta_{i+1} = \frac{d_{i+1}}{d_i}$. We denote $\delta_1 = x > 0$. Hence, we set

$$f_{i+1}(x) = \delta_{i+1}, \tag{7}$$

and from (4), we recursively obtain the inequality

$$0 < \lambda_{i+1}f_{i+1}(x) = \lambda_1x - \mu_i f_i^{-1}(x) + c_{i+1}, \tag{8}$$

where $c_{i+1} = \lambda_i + \mu_{i+1} - \lambda_0 - \mu_1$.

Note that the derivatives of the functions $f_{i+1}(x)$ are strictly greater than zero, and therefore the functions $f_{i+1}(x)$ increase and have exactly one root. The function $f_1(x), x \geq 0$, has the unique root $x_1 = 0$. Through recursion, we obtain from (8) that the function $f_{i+1}(x)$ has a unique positive root, for example, x_{i+1} and $x_{i+1} > x_i, i > 0$.

We have $\delta_k = f_k(x_i)$ and

$$a_k(x_i) > 0, k < i > 1. \tag{9}$$

Let $x_* = \lim_{i \rightarrow \infty} x_i$. From the definition of x_i and the monotonicity of the functions f_i , we have $f_i(x) \geq f_i(x_*) > 0$ for all $x > x_*, i = 1, 2, \dots$. Hence, in the case considered, the set of solutions of $\alpha_i = \alpha_0 > 0$ is given by

$$\delta_i^* = f_i(x) > 0,$$

$$x_* \leq x \leq \frac{\lambda_0 + \mu_1}{\lambda_1} = x^*, i \geq 1.$$

□

Theorem 1 states that the sequence d_k always exists and specifies the interval on which the number x is located, using which one can construct d_k and obtain an exact estimate. Unfortunately, in the case of a countable number of states, knowledge of the interval does not allow one to find this x , and instead, a sequence giving a rough estimate is constructed in other ways. In the case of a finite number of states, this x can be found, and accordingly, the sequence d_k will be obtained. Moreover, it will coincide with an eigenvector for the matrix B .

Consider now a null ergodic situation. Let there exist limits $\lambda^* = \lim_{k \rightarrow \infty} \lambda_k$ and $\lim_{k \rightarrow \infty} \mu_k = \mu^*$ such that $\mu^* < \lambda^*$. Set $d_1 = 1, d_{k+1} = \delta_k d_k$. Consider now the diagonal matrix $D = \text{diag}\{d_i\}$ and the matrix DAD^{-1} :

$$DAD^{-1} = \begin{pmatrix} -\lambda_0 & \delta_1^{-1}\mu_1 & 0 & \dots \\ \delta_1\lambda_0 & -(\lambda_1 + \mu_1) & \delta_2^{-1}\mu_2 & \dots \\ 0 & \delta_2\lambda_1 & -(\lambda_2 + \mu_2) & \dots \\ 0 & 0 & \delta\lambda_2 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{10}$$

Denote the sum of all elements of the k -th coloumn of DAD^{-1} by $-\alpha_k$. One can see that $\delta_1 \in (0, 1)$ implies that $\alpha_0 > 0$. Moreover, if $1 - \delta_1$ is sufficiently small, then there exists $\delta_2 \in (0, 1)$ such that $\alpha_1 > 0$, and so on. On the other hand, if we set $\delta^* = \sqrt{\frac{\mu^*}{\lambda^*}}$, then we have $\alpha^* = (\sqrt{\lambda^*} - \sqrt{\mu^*})^2 > 0$. Hence, for the corresponding δ_k , all α_k are positive for large k . Finally, we have $\beta_* = \inf \alpha_k > 0$, and all $d_k < 1$.

Then, for the corresponding weighted norm, we have, using the logarithmic norm, the following bound:

$$\|\mathbf{p}(t)\|_{1D} = \sum_{i=0}^{\infty} d_i p_i(t) = \|\mathbf{p}(t)\|_{1D} \leq e^{t\gamma(A)_{1D}} \|\mathbf{p}(0)\|_{1D} = e^{-\beta_* t} \sum_{i=0}^{\infty} d_i p_i(0).$$

Hence,

$$d_n p_n(t) \leq e^{-\beta_* t} \sum_{i=0}^{\infty} d_i p_i(0) \leq d^* e^{-\beta_* t} \sum_{i=0}^{\infty} p_i(0) \leq e^{-\beta_* t}, \tag{11}$$

for any n , where $d^* = \sup_k d_k \leq 1$.

This implies the following statement.

Theorem 2. *Let X be a homogeneous BDP with intensities λ_k and μ_k . Then, there exist limits $\lambda^* = \lim_{k \rightarrow \infty} \lambda_k$ and $\lim_{k \rightarrow \infty} \mu_k = \mu^*$ such that $\mu^* < \lambda^*$. Then, the BDP is null ergodic, and there exists a sequence of positive numbers $\{d_i\}$ such that bound (11) holds.*

Remark 1. *All our considerations can be generalized to the case of arbitrary intensity functions $\lambda_k(t)$ and $\mu_k(t)$. Let us consider in particular the most important case of one-periodic intensities. We introduce the notation $\hat{\lambda}_k = \int_0^1 \lambda_k(t) dt$ and $\hat{\mu}_k = \int_0^1 \mu_k(t) dt$. Then, the corresponding results for weakly ergodic and null ergodic situations follow from the existence of $\hat{\lambda}^* = \lim_{k \rightarrow \infty} \hat{\lambda}_k$ and $\lim_{k \rightarrow \infty} \hat{\mu}_k = \hat{\mu}^*$ for $\hat{\lambda}^* < \hat{\mu}^*$ or $\hat{\lambda}^* > \hat{\mu}^*$, respectively, as shown below when considering specific models.*

3. Numerical Examples

Since in real systems service requests usually arrive with variable intensities, the intensities below are taken to be periodic functions. Consider the queue-length process for four specific weakly ergodic models. We suppose that all intensity functions are periodic and that the ‘basic’ intensities are $\lambda(t) = 1 + \sin(2\pi t)$ and $\mu(t) = 1 + \cos(2\pi t)$.

The intensities for the models are given below:

1. Model 1: the client gets scared when he sees the queue and does not join the queue—the rate of arrival decreases.
2. Model 2: the client thinks that something important is being sold when he sees the queue and joins the queue—the rate of arrival increases (a queue for Apple phones).
3. Model 3: with a large number of requests, network performance can drop (Ethernet)—the rate of service decreases.
4. Model 4: the client gets tired of waiting in the queue and leaves it (service intensity + escape) and the server switches to a more productive operating mode with a large number of requests, etc.—the rate of service increases.

The convergence pattern implies that over time the initial state in which the system was ceases to be important and has almost no effect on the behavior of the system. Thus, knowing the rate of convergence, one can find the value of t_* after which all solutions of the system are almost indistinguishable from each other. Therefore, one can construct a limit solution from this t_* using numerical methods.

For the four models, the results obtained can be combined in a table. In the examples, $\delta = \frac{3}{2}$. Taking this into account, the norm of the difference in solutions was obtained.

Since in the examples $X(0) = 1$ and $X(0) = 100$ were taken, $\|p(0) - \pi\|_{1D} \leq \sum_{k=0}^{100} (\frac{3}{2})^k \leq 2 \cdot 10^{18}$.

	Model 1	Model 2	Model 3	Model 4
intensities	$\lambda_k(t) = \frac{\lambda(t)}{k+1},$ $\mu_k(t) = \mu(t)$	$\lambda_k(t) = (1 - \frac{1}{k+2})\lambda(t),$ $\mu_k(t) = \mu(t) + 1$	$\lambda_k(t) = \lambda(t),$ $\mu_k(t) = \mu(t)(1 - \frac{1}{k+1}) + 1$	$\lambda_k(t) = \lambda(t),$ $\mu_k(t) = \frac{\mu(t)}{k+1} + \lambda(t) + 1$
$\gamma^{(B)}_{1D}$ for 'basic' intensities	$-\left(\frac{\mu(t)}{3} - \frac{\lambda(t)}{20}\right)$	$-\left(\frac{1}{3}(\mu(t) + 1) - \frac{1}{2}\lambda(t)\right)$	$-\left(\frac{9}{30}\mu(t) - \frac{1}{2}\lambda(t) + \frac{1}{3}\right)$	$-\left(\frac{1}{3} - \frac{1}{6}\lambda(t)\right)$
$\gamma^{(B)}_{1D}$	$-\frac{17}{60}t$	$-\frac{1}{6}t$	$-\frac{2}{50}t$	$-\frac{1}{6}t$
t_*	180	300	380	300
error on interval $[t_*, t_* + 1]$	0.0003	0.0008	0.0005	0.0008

To plot graphs and for ODE systems, we use standard libraries: Matplotlib and NumPy. We chose numpy because this library allows us to optimize calculations and achieve a higher execution speed for several reasons:

1. NumPy is written primarily in the C programming language, which means that most operations are performed at the machine code level, which provides high performance compared to pure Python code;
2. NumPy arrays are stored compactly and sequentially in RAM, which minimizes overhead and improves processor caching. When data are located close to each other, access to them becomes much faster, especially when processing large amounts of data;
3. Operators and functions inside NumPy are pre-compiled and optimized for various platforms and processor architectures. This avoids the cost of interpreting Python code during program execution.

Here is an example of a code listing that allows a system of differential equations to be solved:

```
import numpy as np
from scipy.integrate import solve_ivp
def function(t, y):
    return fA(t).dot(y)
def main_calc_rk(init_vector, t_list, function):
    return solve_ivp(function, t_list, np.ravel(init_vector),
method='RK45', atol=1e-12, rtol=1e-12)
t_period = [0, t_end + 1]
init_vector = np.zeros((n, 1))
init_vector[0] = 1
solution1 = main_calc_rk(init_vector, t_period, function)
init_vector = np.zeros((n, 1))
init_vector[-1] = 1
solution2 = main_calc_rk(init_vector, t_period, function)
```

- Let $\lambda_k(t) = \frac{\lambda(t)}{k+1}$ and $\mu_k(t) = \mu(t)$; this is a model for the queue with discouragement. Hence, we obtain $\gamma(B(t))_{1D} = -\inf_{k \geq 0} \alpha_k(t)$:

$$\alpha_k = \begin{cases} \mu(t) + \lambda(t) \left(1 - \frac{\delta}{2}\right), & k = 0; \\ \frac{\lambda(t)}{2} + \mu(t) (1 - \delta^{-1}) - \frac{\lambda(t)}{2}, & k = 1; \\ \frac{\lambda(t)}{3} + \mu(t) - \frac{3\lambda(t)}{8} - \frac{2\mu(t)}{3}, & k = 2; \\ \frac{\lambda(t)}{4} + \mu(t) - \frac{3\lambda(t)}{10} - \frac{2\mu(t)}{3}, & k = 3. = \\ \dots & \\ \lambda_n(t) + \mu_{n+1}(t) - \delta\lambda_{n+1}(t) - \delta^{-1}\mu_n(t), & k = n. \\ \dots & \end{cases} \tag{12}$$

$$= \begin{cases} \mu(t) + \frac{\lambda(t)}{4}, & k = 0; \\ \frac{\mu(t)}{3}, & k = 1; \\ \frac{\mu(t)}{3} - \frac{\lambda(t)}{24}, & k = 2; \\ \frac{\mu(t)}{3} - \frac{\lambda(t)}{20}, & k = 3. \\ \dots & \\ \lambda_n(t) + \mu_{n+1}(t) - \delta\lambda_{n+1}(t) - \delta^{-1}\mu(t), & k = n. \\ \dots & \end{cases} \tag{13}$$

$$\gamma(B)_{1D} \leq -\left(\frac{\mu(t)}{3} - \frac{\lambda(t)}{20}\right), \tag{14}$$

$$\|\mathbf{p}(t) - \pi\|_{1D} \leq 2e^{-\frac{17}{60}t} \|\mathbf{p}(0) - \pi\|_{1D}. \tag{15}$$

Detailed considerations (see [15]) show that it is sufficient to limit ourselves to a truncated BDP with states from 0 to 100. The corresponding graphs are presented in Figure 1.

- Let $\lambda_k(t) = (1 - \frac{1}{k+2})\lambda(t)$ and $\mu_k(t) = \mu(t) + 1$; this is a model for the queue with encouragement. Hence, we obtain

$$\alpha_k = \begin{cases} \mu(t) + 1 - \frac{1}{2}\lambda(t), & k = 0; \\ \frac{1}{3}(\mu(t) + 1) - \frac{1}{2}\lambda(t), & k = 1; \\ \dots & \\ \lambda_n(t) + \mu_{n+1}(t) - \delta\lambda_{n+1}(t) - \delta^{-1}\mu_n(t), & k = n. \\ \dots & \end{cases} \tag{16}$$

$$\gamma(B)_{1D} \leq -\left(\frac{1}{3}(\mu(t) + 1) - \frac{1}{2}\lambda(t)\right), \tag{17}$$

$$\|\mathbf{p}(t) - \pi\|_{1D} \leq 2e^{-\frac{1}{6}t} \|\mathbf{p}(0) - \pi\|_{1D}. \tag{18}$$

The corresponding graphs are shown in Figure 2.

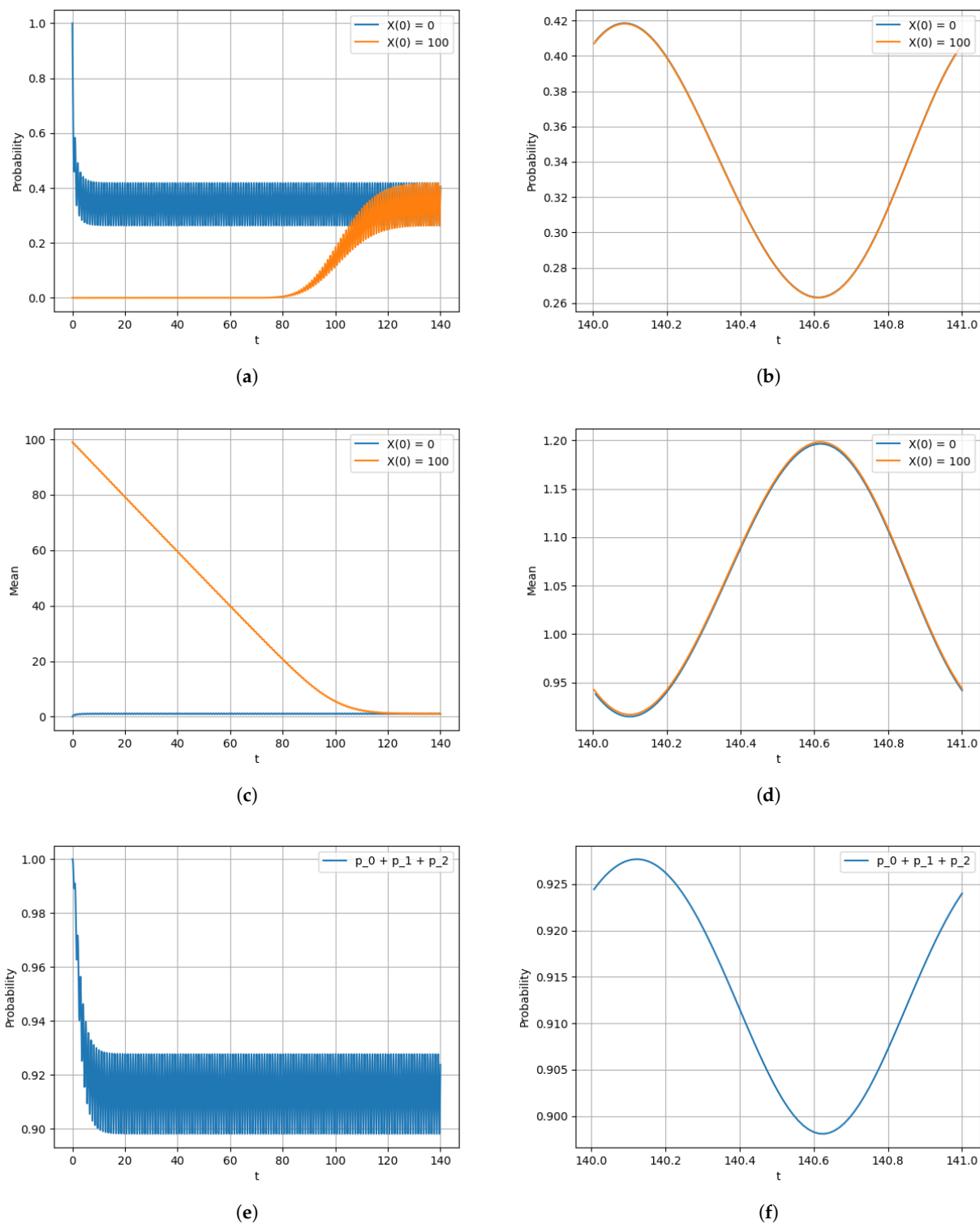


Figure 1. (a) convergence of the probability $p_0(t)$ under the two different initial conditions: $X_{100}(0) = 0$, $X_{100}(0) = 100$; (b) limiting version of the probability $p_0(t)$ under the two different initial conditions: $X_{100}(0) = 0$ and $X_{100}(0) = 100$; (c) the mean $E_N(t, 0)$ on the interval $t \in [0, 140]$, where $N = 100$; (d) the limiting mean on the interval $t \in [140, 141]$, where $N = 100$; (e) the probability $p_0(t) + p_1(t) + p_2(t)$ on the interval $t \in [0, 140]$; (f) limiting version of the probability $p_0(t) + p_1(t) + p_2(t)$ on the interval $t \in [140, 141]$.

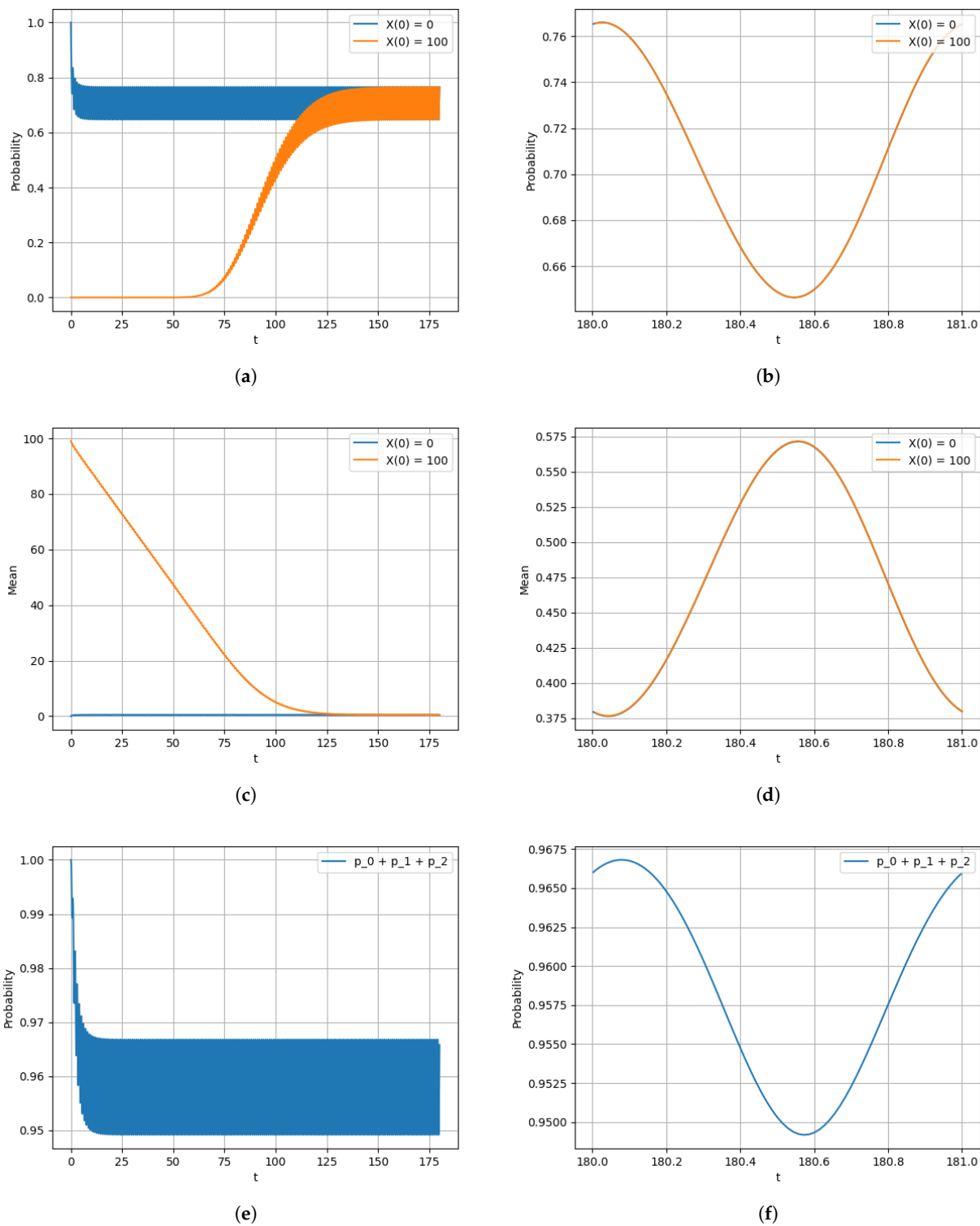


Figure 2. (a) Convergence of the probability $p_0(t)$ under the two different initial conditions: $X_{100}(0) = 0$ and $X_{100}(0) = 100$; (b) limiting version of the probability $p_0(t)$ under the two different initial conditions: $X_{100}(0) = 0$ and $X_{100}(0) = 100$; (c) the mean $E_N(t, 0)$ on the interval $t \in [0, 180]$, where $N = 100$; (d) the limiting mean on the interval $t \in [180, 181]$, where $N = 100$; (e) the probability $p_0(t) + p_1(t) + p_2(t)$ on the interval $t \in [0, 180]$; (f) the limiting version of the probability $p_0(t) + p_1(t) + p_2(t)$ on the interval $t \in [180, 181]$.

3. Let $\lambda_k(t) = \lambda(t)$ and $\mu_k(t) = \mu(t)(1 - \frac{1}{k+1}) + 1$; this is a model for the queue with impatience. Hence, we obtain

$$\alpha_k = \begin{cases} \frac{1}{2}\mu(t) - \frac{1}{2}\lambda(t) + 1, & k = 0; \\ \dots & \\ \frac{9}{30}(\mu(t) + 1) - \frac{1}{2}\lambda(t) + \frac{1}{3}, & k = 4; \\ \dots & \\ \lambda_n(t) + \mu_{n+1}(t) - \delta\lambda_{n+1}(t) - \delta^{-1}\mu_n(t), & k = n. \\ \dots & \end{cases} \quad (19)$$

$$\gamma(B)_{1D} \leq -\left(\frac{9}{30}\mu(t) - \frac{1}{2}\lambda(t) + \frac{1}{3}\right), \quad (20)$$

$$\|\mathbf{p}(t) - \pi\|_{1D} \leq 2e^{-\frac{2}{15}t}\|\mathbf{p}(0) - \pi\|_{1D}. \quad (21)$$

The corresponding graphs are shown in Figure 3.

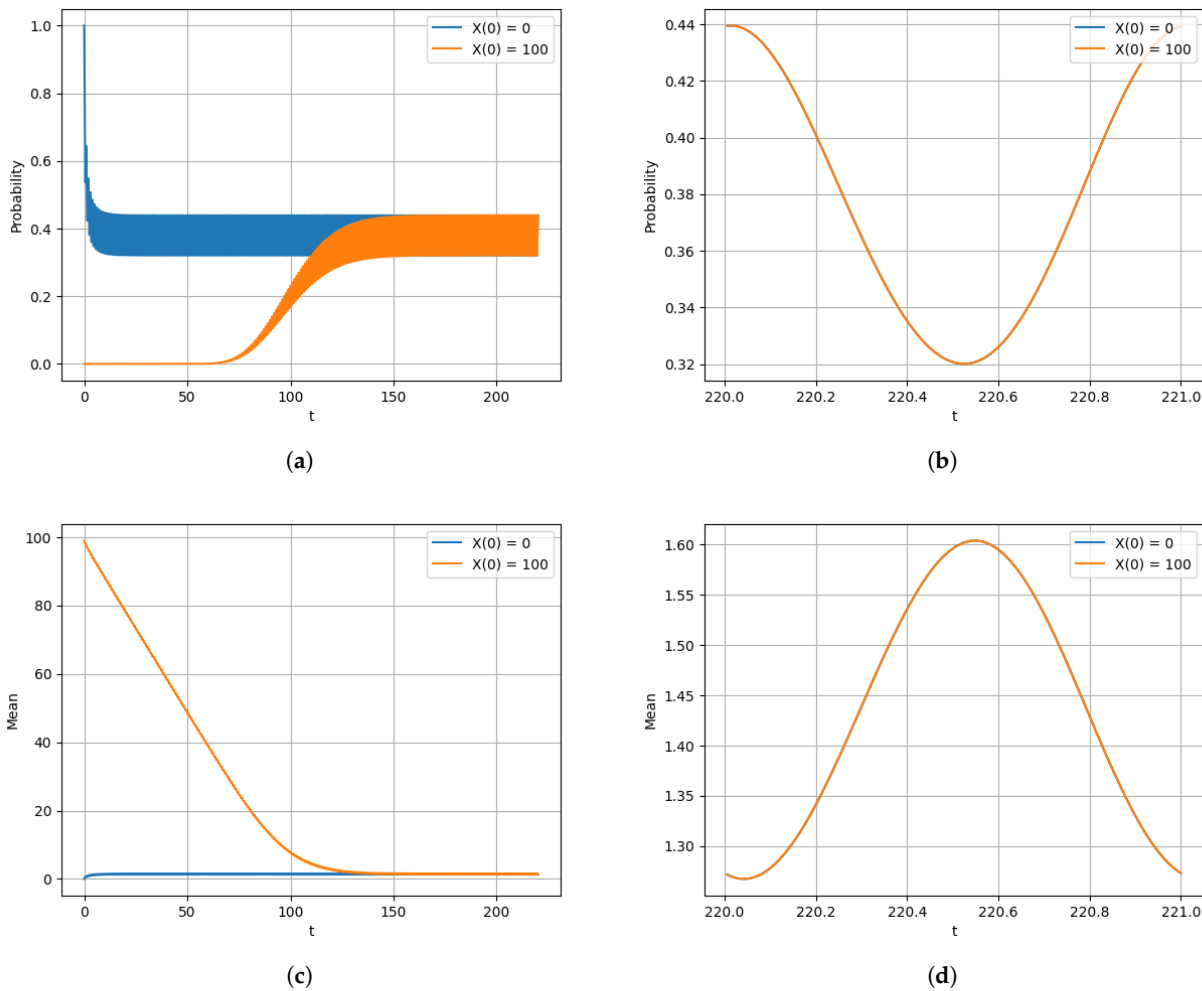


Figure 3. Cont.

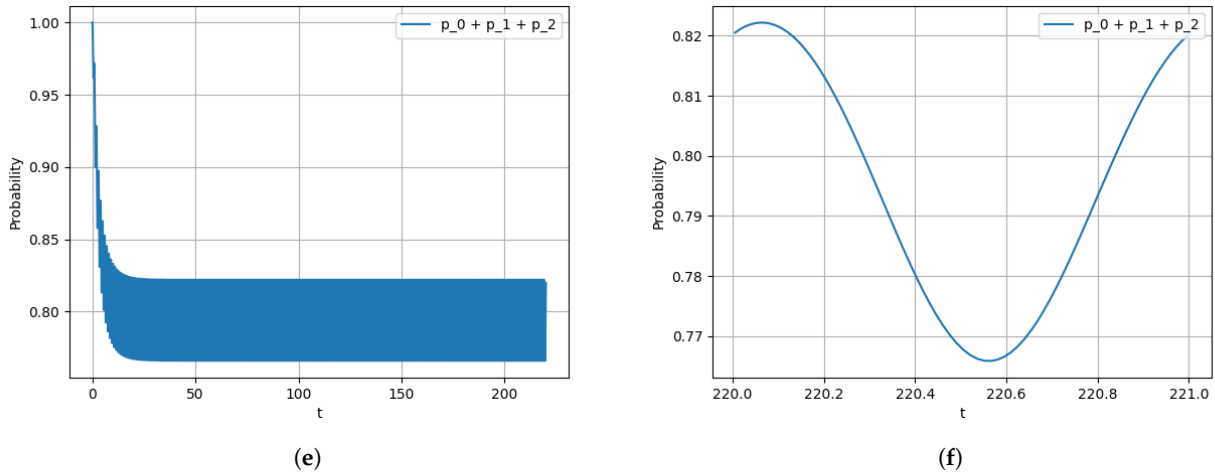


Figure 3. (a) Convergence of the probability $p_0(t)$ under the two different initial conditions: $X_{100}(0) = 0$ and $X_{100}(0) = 100$; (b) limiting version of the probability $p_0(t)$ under the two different initial conditions: $X_{100}(0) = 0$ and $X_{100}(0) = 100$; (c) the mean $E_N(t, 0)$ on the interval $t \in [0, 220]$, where $N = 100$; (d) the limiting mean on the interval $t \in [220, 221]$, where $N = 100$; (e) the probability $p_0(t) + p_1(t) + p_2(t)$ on the interval $t \in [0, 220]$; (f) the limiting version of the probability $p_0(t) + p_1(t) + p_2(t)$ on the interval $t \in [220, 221]$.

- Let $\lambda_k(t) = \lambda(t)$ and $\mu_k(t) = \frac{\mu(t)}{k+1} + \lambda(t) + 1$; this is a model with a server that works less effectively as the queue length increases. Hence, we obtain

$$\alpha_k = \begin{cases} \frac{1}{2}\lambda(t) + \frac{1}{2}\mu(t) + 1, & k = 0; \\ \frac{1}{3} - \frac{1}{6}\lambda(t), & k = 1; \\ \dots & \\ \lambda(t)(1 - \delta) + (\lambda(t) + 1)(1 - \delta^{-1}) + \left(\frac{\mu(t)}{n+2} - \frac{\mu(t)}{\delta(n+1)}\right), & k = n \\ \dots & \end{cases} \quad (22)$$

$$\gamma(B)_{1D} \leq -\left(\frac{1}{3} - \frac{1}{6}\lambda(t)\right), \quad (23)$$

$$\|\mathbf{p}(t) - \pi\|_{1D} \leq 2e^{-\frac{1}{6}t}\|\mathbf{p}(0) - \pi\|_{1D}. \quad (24)$$

The corresponding graphs are shown in Figure 4.

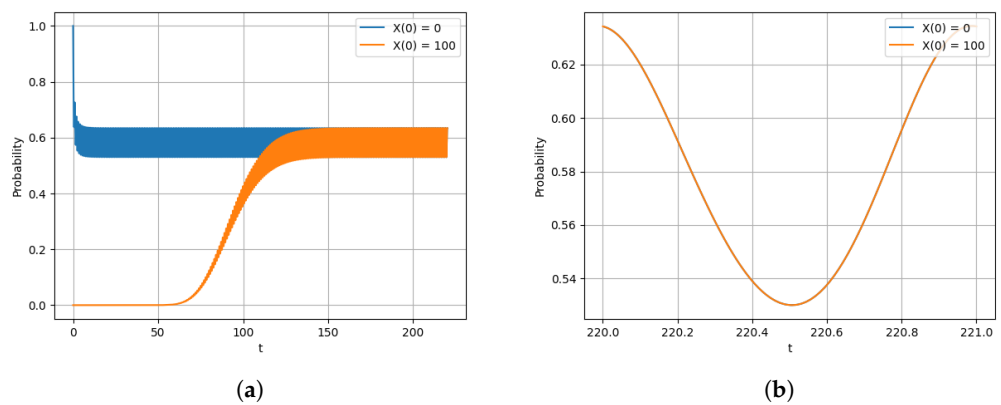


Figure 4. Cont.

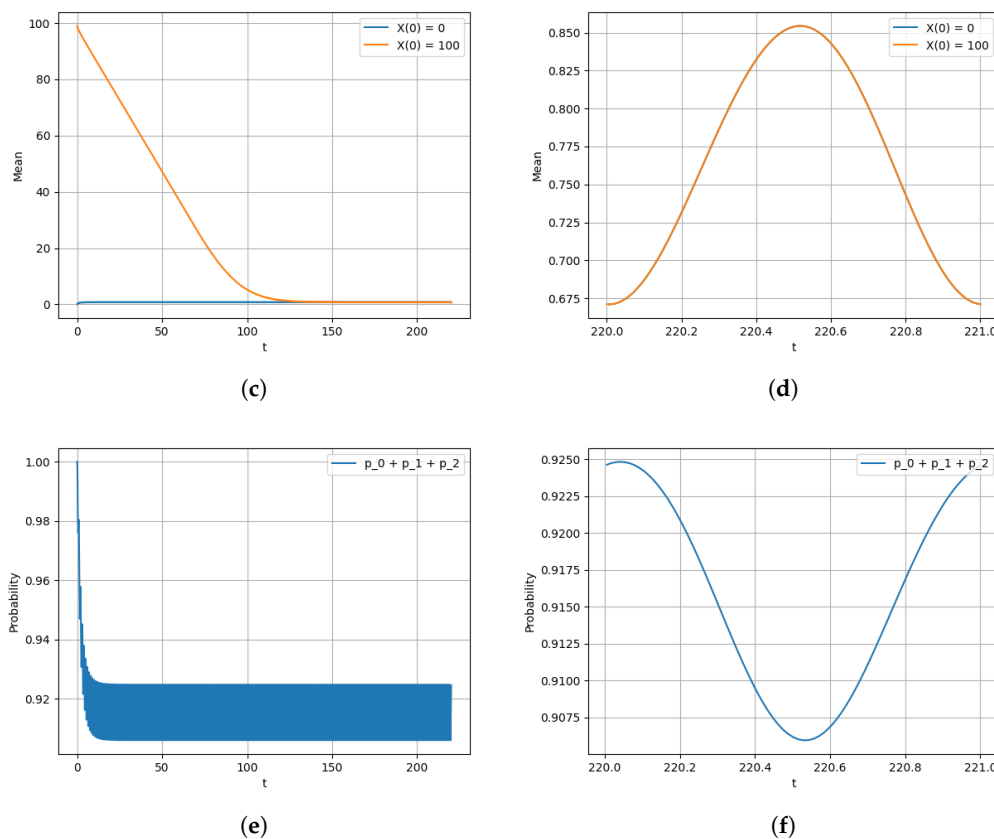


Figure 4. (a) Convergence of the probability $p_0(t)$ under the two different initial conditions: $X_{100}(0) = 0$ and $X_{100}(0) = 100$; (b) limiting version of the probability $p_0(t)$ under the two different initial conditions: $X_{100}(0) = 0$ and $X_{100}(0) = 100$; (c) the mean $E_N(t,0)$ on the interval $t \in [0, 220]$, where $N = 100$; (d) the limiting mean on the interval $t \in [220, 221]$, where $N = 100$; (e) the probability $p_0(t) + p_1(t) + p_2(t)$ on the interval $t \in [0, 220]$; (f) the limiting version of the probability $p_0(t) + p_1(t) + p_2(t)$ on the interval $t \in [220, 221]$.

4. Conclusions

Possible variants of the simplest models of queueing systems of the $M_n(t)/M_n(t)/1$ type are considered. The queue-length process in such systems is described by a BDP, in which the rates of birth (arrivals) and death (servicing) depend on the behavior of customers and the server. For each queue type, the corresponding specific nonstationary system is considered in the weakly ergodic case. Estimates of the rate of convergence to the limiting regime were obtained. The presence of such estimates provides a guaranteed opportunity to compute the limiting characteristics of the system, and therefore, from [15], we can obtain the corresponding estimates of perturbations and approximations via truncations.

Author Contributions: Conceptualization, supervision, A.Z.; methodology, Y.S.; software, validation, visualization I.U.; investigation, writing, I.U., Y.S., A.Z. and J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Natvig, B. On the transient state probabilities for a queueing model where potential customers are discouraged by queue length. *J. Appl. Probab.* **1974**, *11*, 345–354. [CrossRef]
2. Ammar, S.I.; El-Sherbiny, A.A.; El-Shehawy, S.A.; Al-Seedy, R.O. A matrix approach for the transient solution of an $M/M/1/N$ queue with discouraged arrivals and reneging. *Int. J. Comput. Math.* **2012**, *89*, 482–491. [CrossRef]
3. Van Doorn, E.A. The transient state probabilities for a queueing model where potential customers are discouraged by queue length. *J. Appl. Probab.* **1981**, *18*, 499–506. [CrossRef]
4. Reynolds, J.F. The stationary solution of a multiserver queueing model with discouragement. *Oper. Res.* **1968**, *16*, 64–71. [CrossRef]
5. Baccelli, F.; Hebuterne, G. On Queues with Impatient Customers [Research Report] RR-0094. INRIA. 1981. inria-00076467. Available online: <https://inria.hal.science/inria-00076467/document> (accessed on 1 January 2020).
6. Wang, K.; Li, N.; Jiang, Z. Queueing system with impatient customers: A review. In Proceedings of the 2010 IEEE International Conference on Service Operations and Logistics, and Informatics, QingDao, China, 15–17 July 2010; IEEE Xplore: New York, NY, USA, 2010; pp. 82–87. [CrossRef]
7. Sudhesh, R. Transient analysis of a queue with system disasters and customer impatience. *Queueing Syst.* **2010**, *66*, 95–105. [CrossRef]
8. Dudin, A.N.; Dudin, S.A.; Klimenok, V.I.; Dudina, O.S. Stability of queueing systems with impatience, balking and non-persistence of customers. *Mathematics* **2024**, *12*, 2214. [CrossRef]
9. Jouini, O.; Roubos, A. On multiple priority multi-server queues with impatience. *J. Oper. Res. Soc.* **2014**, *65*, 616–632. Available online: <https://www.tandfonline.com/doi/abs/10.1057/jors.2012.153> (accessed on 1 January 2020). [CrossRef]
10. Puha, A.L.; Ward, A.R. Scheduling an overloaded multiclass many-server queue with impatient customers. In *Operations Research & Management Science in the Age of Analytics*; INFORMS: Catonsville, MD, USA, 2019; pp. 189–217. Available online: <https://pubsonline.informs.org/doi/abs/10.1287/educ.2019.0196> (accessed on 1 January 2020).
11. Dong, J.; Ibrahim, R. SRPT scheduling discipline in many-server queues with impatient customers. *Manag. Sci.* **2021**, *67*, 7708–7718. Available online: <https://pubsonline.informs.org/doi/abs/10.1287/mnsc.2021.4110> (accessed on 1 January 2020). [CrossRef]
12. Som, B.K. A Markovian feedback queueing model with encouraged arrivals and customers impatience. In Proceedings of the 2nd International Conference on Evidence Based Management, BITS, Pilani, India, 17–18 March 2017. Available online: https://scholar.google.com/scholar?lookup=0&q=A+Markovian+feedback+queueing+model+with+encouraged+arrivals+and+customers+impatience.&hl=ru&as_sdt=0,5 (accessed on 1 January 2020).
13. Som, B.K.; Seth, S. $M/M/c/N$ queueing systems with encouraged arrivals, reneging, retention and feedback customers. *Yugosl. J. Oper. Res.* **2018**, *28*, 333–344. <https://doi.org/0.2298/YJOR170620006S> [CrossRef]
14. Ingolfsson, A.; Akhmetshina, E.; Budge, S.; Li, Y.; Wu, X. A survey and experimental comparison of service-level approximation methods for nonstationary $M(t)/M/s(t)$ queueing systems with exhaustive discipline. *INFORMS J. Comput.* **2007**, *19*, 149–312. [CrossRef]
15. Zeifman A.; Satin Y.; Korolev V.; Usov I. Continuous-time Markov chains and models: Bounding via forward Kolmogorov system. In *Handbook of Statistics*; Elsevier: Amsterdam, The Netherlands, 2025; Volume 52, pp. 179–210. Available online: <https://www.sciencedirect.com/science/article/abs/pii/S0169716125000069> (accessed on 1 June 2025).
16. Granovsky, B.L.; Zeifman, A.I. The decay function of nonhomogeneous birth-death processes, with application to mean-field models. *Stoch. Proc. Appl.* **1997**, *72*, 105–120. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.