Chapter 6

Single Server Queueing Model with Feedback

6.1 Introduction

The studied $M/M/1$ service model with feedback can be visualized as scheduling mechanisms in multi-access service systems like time-sharing computer systems. The single-processor gives a quantum of time in random to each specific process. A process once started is allowed to run either until it has exhausted its quantum or until it initiates a transfer to a peripheral device, or until an interrupt generated by some other process occurs. If the allotted random quantum is exhausted, the process is assigned a longer quantum and rejoins the queue. If the process initiates a transfer, its quantum remains unaltered, and it rejoins the queue. If an externally generated interrupt occurs, the interrupt is serviced. Servicing the interrupt may free some other process already in the queue, in which case that process may be preferentially restarted. Queue with feedback is defined as after being served each customer either immediately rejoins the queue again randomly with some probability or departs permanently with compliment probability.

Considerable efforts have been directed to drive docile and abstractly easy format for the transient state probabilities for a single server queue. In decade of 50's, significant contributions have been observed to drive time-dependent solution of queue-based service system with single server and continuous parameters using the mathematical concepts of generating functions, Laplace transform theory, complex analysis or combinatorics, spectral method et cetera (*cf.* [50], [24], [159], [37]). Kawamura [124] analyzed the results on how fast the transient state probabilities and the

[0]The partial content of this chapter has been published in
C. Shekhar, A. Kumar, and **S. Varshney**, "Modified Bessel Series Solution of the Single Server Queueing Model with Feedback", *International Journal of Computing Science and Mathematics*, vol. 10, no.3. pp. 313-326, 2019, (INDERSCIENCE) (`https://www.inderscience.com/info/inarticle.php?artid=101101`).

mean value of the customers in the system tended to their equilibrium limits when these existed. Parthasarathy [203] derived transient state probabilities of the classical $M/M/1$ queueing model using modified Bessel function. Krinik [145] presented a Taylor-series method for determining the transient probabilities of the classical single server queue. Leguesdron et al. [165] proposed a new method based on the uniformization technique of the Markov chain and the inversion of generating functions. Sharma and Tarabia [224] obtained a series form for the transient state probabilities of the single server Markovian queue with finite waiting space and derived a recurrence relation for the coefficients. Tarabia [257] proposed a simple alternative approach using Chebychev's polynomial for the transient analysis of finite capacity single server queue.

Takács [248] introduced the concept of feedback in queueing modeling with a single server, and the author determined the distribution of the queue size, as well as the Laplace-Stieltjes transform for a stationary process and the first two moments of the distribution function of the total time spent in the system by a customer. Chan [38] used generating function to obtain a necessary and sufficient condition for the existence of statistical equilibrium of the multi-server queue with feedback in the information processing unit. Using Markov decision theory and convexity arguments, [49] considered a system of two coupled queues where a packet after being served in one queue may be fed into the other queue or may leave the system. Using matrix-geometric method, [19] computed state probabilities of the service system at call center considering retrial and feedback facility for calls. Transient state queue length distribution and Laplace transform of their generating function were derived by [239] in feedback queue with correlated departures. Exact and approximate methods to calculate performance characteristics of the system were developed by [194] for the multi-channel queueing model with feedback which occurred as returning a part of serviced calls to get a new service.

Even nowadays, there is continuing interest to develop new insights and methods for a better understanding of transient behavior in the classical single server queue-based service system with some controllable arrival and controllable service policy. Asymptotic behavior of the transient state probability of the server being idle and mean queue size was discussed by [148] in the $M/M/1$ queueing model with the possibility of catastrophes at the service station. Using modified Bessel function, [7] derived a transient solution for infinite server queues with Poisson arrivals and exponential service times with time-dependent arrival and service rate. Kumar and Madheswari [149] obtained a transient solution for the system size in the $M/M/1$ queue with the possibility of catastrophes and server failure. Griffiths et al. [84]

presented an advanced generalization of the modified Bessel function and its generating function. Later, [166] used the results derived by [84] as a generalization of the modified Bessel function to determine transient-state probabilities of the single-server queueing model with Erlangian service in k phases. Al-Seedy et al. [8] used the technique of generating function to evolve the transient-state probabilities for a multi-server system incorporating the impatience behavior of balking with fixed probability and negative exponentially distributed reneging. Ammar et al. [11] illustrated computable matrix technique that can be used to derive an explicit expression for the transient-state probabilities of a finite waiting space single-server queue having discouraged arrivals and reneging. Kalidass and Ramanath [121] obtained explicit expressions for the time-dependent probabilities of the $M/M/1$ queue with server vacations under a multiple vacation scheme.

The practical applications of our studied service model can be observed in the time-sharing interactive computer system. The interactive computer system consists of the user requests and then receives service from the computer system. The events usually forming an interaction are the user's thinking, typing at his remote console, waiting for a response from the computer system, and finally, watching the output. These interactions are repeated until the user finds the desired output. The number of interactions depends on the contents of a job, which is processed by the computer system and on the goodness of the program, which is processed by the user in each interaction. Since this number fluctuates stochastically, it may be considered as a random variable.

The objective of choosing the classical service model and using the methodology for deriving an explicit expression for the state of the service system is threefold. (1) Exponential distribution allows for a very simple portrayal of the state of the service system at time t, namely the number of customers in the service system. Since the exponential distribution is memoryless, neither we do have to recollect when the previous customer arrived, nor we have to register when the earlier customer entered service. (2) The classical service model seems like the foundation for the extended study of the complex service systems for developing a better service system. (3) The derived method is simple in the procedure and easy to understand and use only some standard mathematical concepts.

The organization of the chapter is as follows: In section 6.2, we describe the service model using some assumptions and notations and formulate the governing equations. In section 6.3, we present a simple procedure to determine transient-state probabilities of the $M/M/1$ queue-based service system with feedback using a modified Bessel function. In section 6.4, we derived the closed-form of some measure of effectiveness. In section 6.5, we summarize the sensitivity of stationary

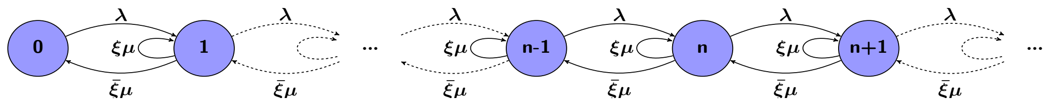

Figure 6.1: State Transition diagram

process for continuous parameters. Finally, in section 6.6, we remark the conclusion and discuss the future scope.

6.2 Model Description

In this chapter, we analyze the service model with exponential inter-arrival times with mean $1/\lambda$, exponential service times with mean $1/\mu$, and a single server. The served customers may rejoin the service system for completing the unsatisfactory service with probability ξ or may leave the service system with probability $\bar{\xi} = (1-\xi)$. Let the random process $N(t)$ be the number of the customers waiting in the queue and being served by the server at time t. Suppose $P_n(t) = Pr[N(t) = n]$ *i.e.* probability that there are n; $n = 0,1,2,3,...$ customers in the service system at the time t is the transient probability function.

Based on the memoryless property, the following set of Chapman-Kolmogorov differential-difference infinite equations for the transient-state probabilities $P_n(t)$ is obtained as follows

$$\begin{aligned}\frac{dP_0(t)}{dt} &= -\lambda P_0(t) + \bar{\xi}\mu P_1(t)\\ \frac{dP_n(t)}{dt} &= -(\lambda+\mu)P_n(t) + \lambda P_{n-1}(t) + \xi\mu P_n(t) + \bar{\xi}\mu P_{n+1}(t), n = 1,2,3,...\end{aligned} \tag{6.1}$$

We assume that initially there are $m\,(m \geq 0)$ customers in the service system *i.e.*

$$P_n(0) = \begin{cases} 1; & n = m \\ 0; & n \neq m \end{cases} ; n \geq 0 \tag{6.2}$$

6.3 The Transient Solution

We develop simple and direct approach to compute an explicit expression for transient-state probabilities $P_n(t); n = 0,1,2,...$ from set of birth and death eq^n's(6.1). Let us define

$$F_n(t) = \begin{cases} exp\left\{(\bar{\xi}\mu+\lambda)t\right\}\left[\bar{\xi}\mu P_n(t)-\lambda P_{n-1}(t)\right]; & n=1,2,3,... \\ 0; & \text{otherwise} \end{cases} \tag{6.3}$$

and consider its generating function as

$$K(s,t) = \sum_{n=-\infty}^{\infty} F_n(t)s^n \tag{6.4}$$

Using the eq^n's(6.1) & (6.3) and differentiating eq^n(6.4) with respect to t, we obtain

$$\begin{aligned} \frac{\partial K(s,t)}{\partial t} &= \left(\lambda s+\frac{\bar{\xi}\mu}{s}\right)K(s,t)-\bar{\xi}\mu F_1(t) \\ K(s,0) &= s^m\{\bar{\xi}\mu(1-\delta_{0m})-\lambda s\} \end{aligned} \tag{6.5}$$

where δ_{0m} is the kronecker delta defined as

$$\delta_{0m} = \begin{cases} 0; & m \neq 0 \\ 1; & m = 0 \end{cases} \tag{6.6}$$

Since eq^n(6.5) is linear differential equation in $K(s,t)$, the solution is given by

$$K(s,t)exp\left\{-\left(\lambda s+\frac{\bar{\xi}\mu}{s}\right)t\right\} = -\int exp\left\{-\left(\lambda s+\frac{\bar{\xi}\mu}{s}\right)y\right\}\left(\bar{\xi}\mu\right)F_1(y)dy+g(s) \tag{6.7}$$

Using initial condition, we get $g(s)=K(s,0)$. Hence,

$$K(s,t) = K(s,0)exp\left\{\left(\lambda s+\frac{\bar{\xi}\mu}{s}\right)t\right\}-\bar{\xi}\mu\int_0^t F_1(y)exp\left\{\left(\lambda s+\frac{\bar{\xi}\mu}{s}\right)(t-y)\right\}dy \tag{6.8}$$

It is well known from the literature of Bessels function (Griffiths et al. [85]) that for $\alpha = 2\sqrt{(\lambda\bar{\xi}\mu)}$ and $\beta = \sqrt{\left(\frac{\lambda}{\bar{\xi}\mu}\right)}$, the generating function is given by

$$exp\left\{\left(\lambda s+\frac{\bar{\xi}\mu}{s}\right)t\right\} = \sum_{n=-\infty}^{\infty}(\beta s)^n I_n(\alpha t) \tag{6.9}$$

where $I_n(\alpha,t)$ is a modified Bessel function. On comparing the coefficient of power of s, *i.e* s^n; $n = 1,2,3,...$ on both side of eq^n(6.8), we get

$$\begin{aligned}\beta^{-n}F_n(t) = \bar{\xi}\mu(1-\delta_{0m})\beta^{-m}I_{n-m}(\alpha t) - \lambda\beta^{-m-1}I_{n-m-1}(\alpha t)\\ -\bar{\xi}\mu\int_0^t F_1(y)I_n(\alpha(t-y))dy\end{aligned} \tag{6.10}$$

Since from eq^n(6.3) $F_n(t) = 0$ for $n < 0$, the eq^n(6.10) holds for $n = -1,-2,-3,...$ with the left-hand side replaced by zero. Using the recurrence relation $I_{-r} = I_r$ for $r = 1,2,3,...$ we have

$$\begin{aligned}\bar{\xi}\mu\int_0^t F_1(y)I_n(\alpha(t-y))dy = \bar{\xi}\mu(1-\delta_{0m})\beta^{-m}I_{n+m}(\alpha t)\\ -\lambda\beta^{-m-1}I_{n+m+1}(\alpha t);\ n = 1,2,3,\ldots\end{aligned} \tag{6.11}$$

Form eq^n's(6.10) and (6.11), we have

$$\begin{aligned}F_n(t) = \bar{\xi}\mu\beta^{n-m}(1-\delta_{0m})\left[I_{n-m}(\alpha t) - I_{n+m}(\alpha t)\right]\\ +\lambda\beta^{n-m-1}\left[I_{n+m+1}(\alpha t) - I_{n-m-1}(\alpha t)\right];\ n \geq 1\end{aligned} \tag{6.12}$$

Hence, from eq^n(6.3) and iteration method, we have explicit expression for queue size distribution as follows

$$P_0(t) = \int\limits_0^t F_1(y)exp\left\{-(\lambda+\bar{\xi}\mu)y\right\}dy + \delta_{0m} \tag{6.13}$$

and

$$P_n(t) = \frac{exp\left\{-(\lambda+\bar{\xi}\mu)(t)\right\}}{\bar{\xi}\mu}\sum_{i=1}^{n}F_i(t)\left(\frac{\lambda}{\bar{\xi}\mu}\right)^{n-i} + \left(\frac{\lambda}{\bar{\xi}\mu}\right)^{n}P_0(t) \tag{6.14}$$

Besides the modified Bessel function of the second kind used above to compute the transient-state probabilities, we can employ another kind of Bessel function also with some significant modification to obtain similar results. We prefer the second kind since the resultant expression in the computation of the solution is much resembled with a standard expression of the Bessel function of the second kind. We can further use some other special functions also [241]. Some mathematical approaches like Laplace transform, generating function, etc. for computing the transient-state probabilities can also be used, but it requires Rouchés theorem and have to identify zeros of complicated algebraic equation [74].

6.4 Measure of Effectiveness

In this section, we derive an expression of expected number of customers in the service system $E_N(t)$ in terms of derived queue size distribution in previous section 6.3. Let us consider $\rho = \left(\frac{\lambda}{\bar{\xi}\mu}\right)$. From eq^n(6.14), we have

$$P_n(t) = \frac{exp\{-(\lambda+\bar{\xi}\mu)t\}}{\bar{\xi}\mu}\sum_{i=1}^{n} F_i(t)\rho^{(n-i)} + \rho^n P_0(t) \tag{6.15}$$

We know

$$\begin{aligned}
E_N(t) = E(N(t)) &= \sum_{n=1}^{\infty} nP_n(t)\\
&= \left[1.P_1(t) + 2.P_(t) + 3.P_3(t) + ...\right]\\
&= \left\{\left(\frac{exp\{-(\lambda+\bar{\xi}\mu)t\}}{\bar{\xi}\mu}\right)F_1(t) + \rho P_0(t)\right\}\\
&\quad + 2\left\{\left(\frac{exp\{-(\lambda+\bar{\xi}\mu)t\}}{\bar{\xi}\mu}\right)[\rho F_1(t) + F_2(t)] + \rho^2 P_0(t)\right\} + ...
\end{aligned}$$

$$\begin{aligned}
= \frac{exp\{-(\lambda+\bar{\xi}\mu)t\}}{\bar{\xi}\mu}\left[\frac{1}{(1-\rho)^2}F_1(t) + \left\{\frac{(2-\rho)}{(1-\rho)^2}\right\}F_2(t) + ...\right]\\
+ \frac{\rho}{(1-\rho)^2}P_0(t)
\end{aligned}$$

$$\begin{aligned}
= \frac{exp\{-(\lambda+\bar{\xi}\mu)t\}}{\bar{\xi}\mu}\left[\frac{1}{(1-\rho)^2}\sum_{i=1}^{\infty}F_i(t) + \frac{1}{(1-\rho)}\sum_{i=1}^{\infty}(i-1)F_i(t)\right]\\
+ \frac{\rho}{(1-\rho)^2}P_0(t)
\end{aligned}$$

Hence,

$$E_N(t) = \frac{1}{(1-\rho)^2}\left[\frac{exp\{-(\lambda+\bar{\xi}\mu)t\}}{\bar{\xi}\mu}\left(\sum_{i=1}^{\infty}F_i(t) + (1-\rho)\sum_{i=1}^{\infty}(i-1)F_i(t)\right) + \rho.P_0(t)\right] \tag{6.16}$$

where $F_i(t)$ is derived in eq^n(6.12).

Now for sufficiently large value of t, *i.e.* $t \to \infty$, We get

$$|E_N(t) - L| < \varepsilon \tag{6.17}$$

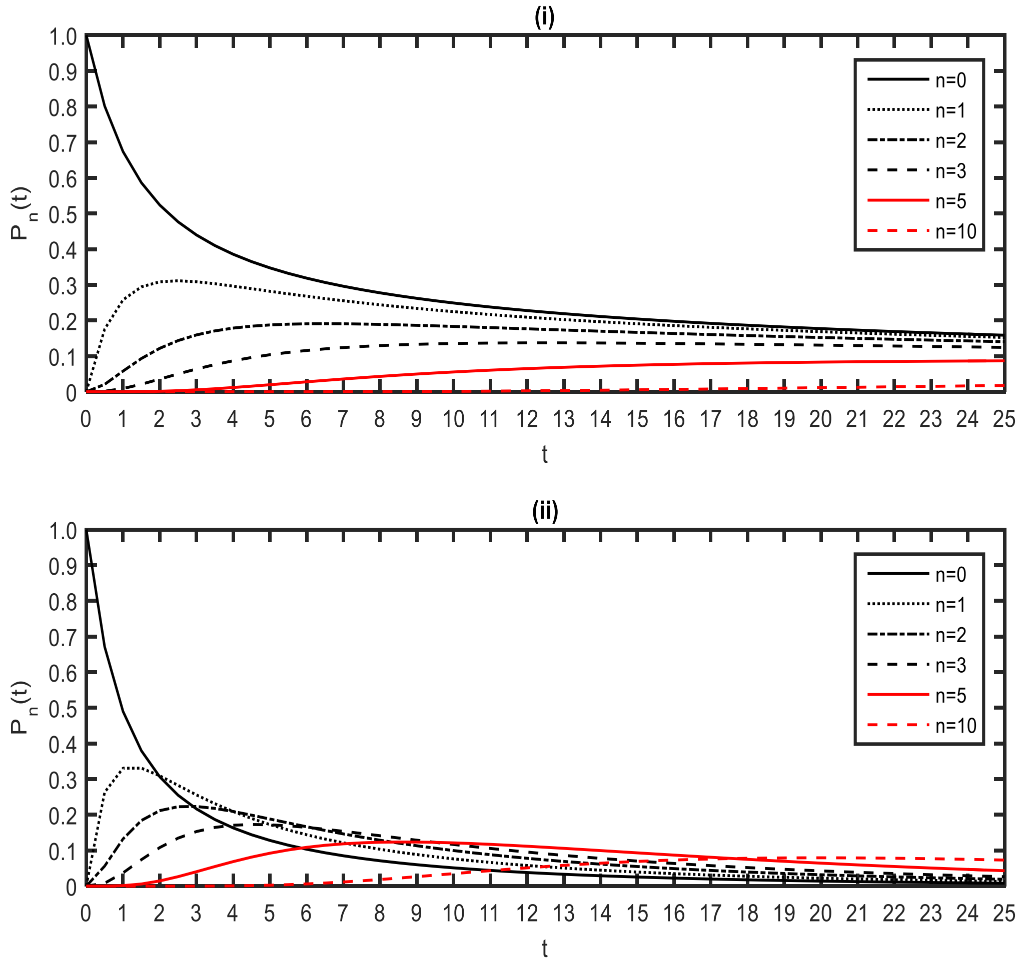

Figure 6.2: Queue size distribution $P_n(t)$ wrt t for different n

where L is expected number of customers in the service system in steady-state condition and given by

$$L = \frac{\rho}{(1-\rho)} \tag{6.18}$$

6.5 Numerical Results

We develop the simple and direct approach for computing an elegant explicit expression for transient-state probabilities of the classical $M/M/1$ queue with feedback. We have also computed the expression for the expected number of customers in the service system in terms of the derived state probabilities. To validate the findings, we illustrate some numerical experiments. For that purpose, we assume that there is no customer in the service system initially. The results are depicted in Figs. 6.2–6.5 and Tables 6.1–6.3.

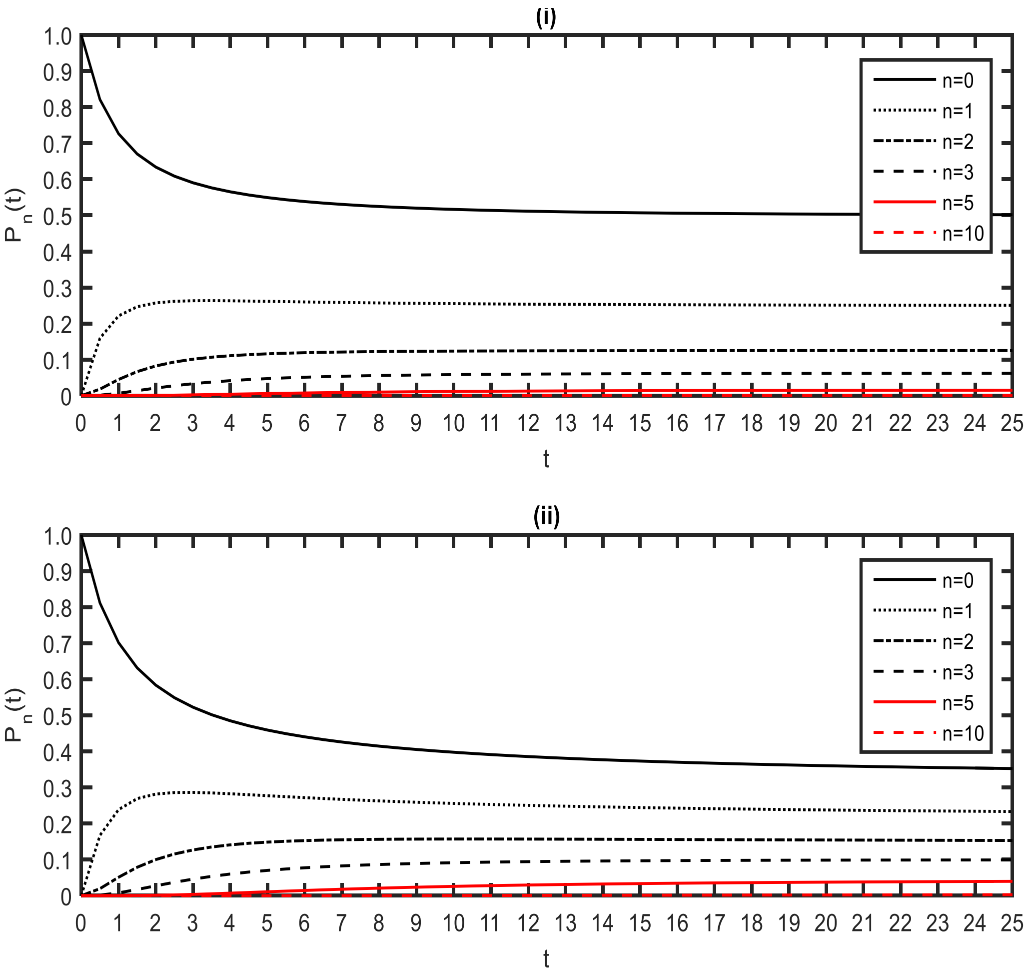

Figure 6.3: Queue size distribution $P_n(t)$ wrt t for different n

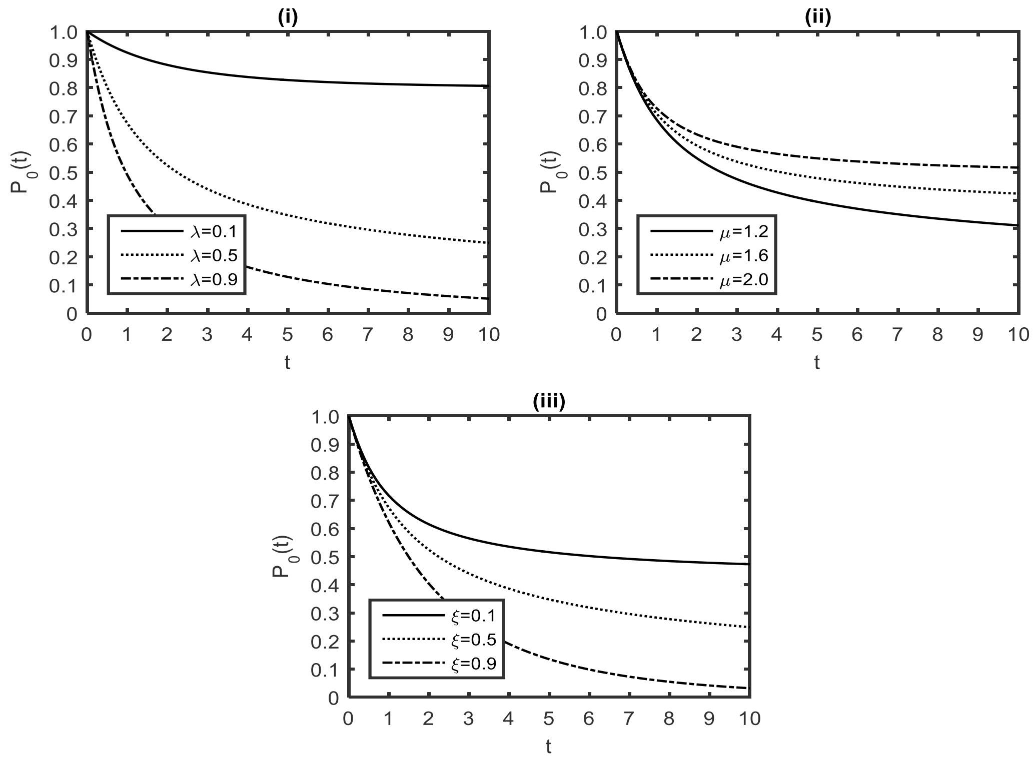

Figure 6.4: Probability that there is no customer $P_0(t)$ wrt t for different parameters

We fix $\lambda = 0.5$, $\mu = 1.0$, $\xi = 0.5$ in Fig. 6.2(i) and $\lambda = 0.9$, $\mu = 1.0$, $\xi = 0.5$ in Fig. 6.2(ii) whereas we take $\lambda = 0.5$, $\mu = 2.0$, $\xi = 0.5$ in Fig. 6.3(i) and $\lambda = 0.5$, $\mu = 1.0$, $\xi = 0.25$ in Fig. 6.3(ii). Figs. 6.2 and 6.3 illustrate how transient-state probabilities $P_n(t)$ for $n = 0,1,2,3,5,10$ very in range from 0 to 25 unit time. It is observed from all the figures that state probabilities are tending to a constant value which leads to a steady-state. It is also prompt from Fig. 6.2(ii) that probability of more customers with time increases for high λ. By taking, it is clear that the probability of fewer customers increases with the increase of service rate μ and probability ξ.

In Fig. 6.4, we demonstrate the variability of the probability of no customer in the service system $P_0(t)$ with time and system parameters for $\lambda = 0.5$, $\mu = 1.0$, $\xi = 0.5$. As time spends more, $P_0(t)$ decreases to a stable value. It is also noticed that its value gets decreased with the increment in the value of arrival rate λ and the probability of feedback ξ and increases with service rate μ. Figs. 6.5(i) and (ii) also comprise the variability of $P_0(t)$ for arrival rate and service rate for the different values of probability of feedback for $\lambda = 0.5$ and $\mu = 1.0$ respectively. The result also demonstrates that $P_0(t)$ decreases with an increase of λ and ξ and increases with μ. The red line illustrates the variation of the $P_0(t)$ without feedback policy, which resembles the results of [203].

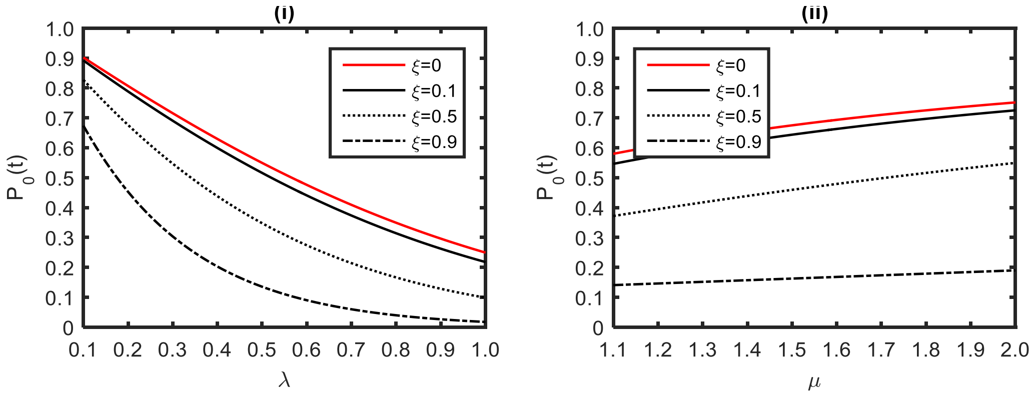

Figure 6.5: Probability that there is no customer $P_0(t)$ wrt parameters for different ξ

Table 6.1: Expected number of customers in the service system wrt λ

λ	t						
	0.10	0.25	0.50	1.00	2.00	5.00	10.00
0.05	0.0049	0.0120	0.0230	0.0433	0.0758	0.1347	0.1744
0.15	0.0148	0.0361	0.0698	0.1305	0.2311	0.4327	0.6178
0.25	0.0246	0.0603	0.1164	0.2184	0.3907	0.7643	1.1775
0.40	0.0394	0.0965	0.1866	0.3515	0.6372	1.3130	2.1964
0.55	0.2557	0.4750	0.6067	0.3812	1.0259	2.0537	3.3767

Table 6.2: Expected number of customers in the service system wrt μ

μ	t						
	0.10	0.25	0.50	1.00	2.00	5.00	10.00
1.0	0.0488	0.1178	0.2232	0.4065	0.7018	1.3293	2.0549
1.5	0.0482	0.1144	0.2113	0.3683	0.5947	0.9876	1.3225
2.0	0.0476	0.1111	0.2002	0.3347	0.5081	0.7511	0.8959
3.0	0.0465	0.1050	0.1802	0.2792	0.3807	0.4720	0.4961
5.0	0.0443	0.0940	0.1478	0.2016	0.2364	0.2494	0.2500

Table 6.3: Expected number of customers in the service system wrt ξ

ξ	t						
	0.10	0.25	0.50	1.00	2.00	5.00	10.00
0	0.0485	0.1164	0.2183	0.3906	0.6561	1.1774	1.7151
0.1	0.0486	0.1172	0.2212	0.4000	0.6830	1.2659	1.9105
0.2	0.0488	0.1180	0.2242	0.4098	0.7114	1.3624	2.1314
0.3	0.0489	0.1189	0.2272	0.4198	0.7412	1.4676	2.3800
0.4	0.0491	0.1197	0.2303	0.4302	0.7726	1.5820	2.6584

In Tables 6.1–6.3, we tabulate the variation of expected number of customers in the service system $E_N(t)$ with respect to time t for a different set of arrival rate (λ), service rate (μ) and probability of feedback (ξ) respectively by fixing $\lambda = 0.5, \mu = 0.6, \xi = 0.5$. Tables 6.1–6.3 show how the expected number of customers in the service system $E_N(t)$ increases with respect to time t. Table 6.1 illustrates increasing behavior of $E_N(t)$ with increment in the value of λ. Table 6.2 shows $E_N(t)$ decreases for higher value of μ which is obvious result of any service system. Table 6.3 depicts that the value of $E_N(t)$ also increases for high value of probability of feedback. These results shows how feedback policy in any queue-based service system makes the system better at the cost of some extra queue size.

6.6 Conclusion and Future Scope

In this chapter, we propose a simple and direct solution technique for computing an explicit expression for transient-state probabilities of the classical $M/M/1$ queue-based service system with feedback. We derive an elegant expression for $P_n(t)$, the equilibrium distribution of the state $N(t)$ using modified Bessel function. We have also illustrated the limiting value of state probabilities and the expected number of customers in the service system analytically and numerically. Moreover, numerical investigations show how parameters λ, μ, and ξ affect the measures of effectiveness significantly. We can extend our approach for more controllable service policy, finite capacity, non-Markovian queues, etc. [8], [121], [85].