

**RAIRO Operations Research**

RAIRO Oper. Res. **38** (2004) 195–213

DOI: 10.1051/ro:2004020

## ENTROPY MAXIMIZATION AND THE BUSY PERIOD OF SOME SINGLE-SERVER VACATION MODELS

JESUS R. ARTALEJO<sup>1</sup> AND MARIA J. LOPEZ-HERRERO<sup>2</sup>

Communicated by E. Gelenbe

**Abstract.** In this paper, information theoretic methodology for system modeling is applied to investigate the probability density function of the busy period in  $M/G/1$  vacation models operating under the  $N$ -,  $T$ - and  $D$ -policies. The information about the density function is limited to a few mean value constraints (usually the first moments). By using the maximum entropy methodology one obtains the least biased probability density function satisfying the system's constraints. The analysis of the three controllable  $M/G/1$  queueing models provides a parallel numerical study of the solution obtained *via* the maximum entropy approach *versus* “classical” solutions. The maximum entropy analysis of a continuous system descriptor (like the busy period) enriches the current body of literature which, in most cases, reduces to discrete queueing measures (such as the number of customers in the system).

**Keywords.** Busy period analysis, maximum entropy methodology,  $M/G/1$  vacation models, numerical inversion.

### 1. INTRODUCTION

A significant part of the results of the “classical” queueing theory are obtained by assuming that the random variability of the inter-arrivals periods and service times can be modelled by probability distributions. The queueing literature shows the existence of a rich variety of exact, approximate and asymptotic methods

---

<sup>1</sup> Department of Statistics and Operations Research, Faculty of Mathematics, Complutense University of Madrid, Madrid 28040, Spain; e-mail: [jesus\\_artalejo@mat.ucm.es](mailto:jesus_artalejo@mat.ucm.es)

<sup>2</sup> School of Statistics, Complutense University of Madrid, Madrid 28040, Spain; e-mail: [lherrero@estad.ucm.es](mailto:lherrero@estad.ucm.es)

for solving the mathematical equations describing the system behavior. Some “classical” and well-known techniques include the general framework of birth-and-death processes and methods of solution for non-Markovian stochastic processes (such as embedded Markov chains or the method of supplementary variables). Nevertheless, it is clear that complex queueing systems arising in real computer networks and teletraffic theory often do not match assumptions made by “classical” queueing modeling. In this context, the use of information theoretic techniques based on the principle of maximum entropy (PME) and the principle of minimum cross-entropy (if a prior distribution is available) provides an elegant alternative to the “classical” stochastic queueing models.

The queueing literature shows that during the last decades there has been an increasing interest in the development of queueing models based on information theoretic methods. We refer the reader to the papers by Shore [20] and El-Affendi and Kouvatso [7] who obtained information theoretic approximation of the most basic queueing models, and to the survey paper by Kouvatso [16]. Some recent applications include the paper of Tadj and Hamdi [21] in the context of single server queues with quorum, and Wang *et al.* [25] who present a maximum entropy analysis for the number of customers in the  $M/G/1$  queue with  $N$ -policy. A novel reader having a first approach to the literature could feed the idea that maximum entropy solutions only provide a reasonable approximation to the true (but complex) queueing system modelled by “classical” techniques. Such interpretation of the information theoretic techniques is poor and trivial. The aim of the PME is to provide a self-contained method of inference for estimating uniquely an unknown probability distribution [13, 16, 20]. The maximum entropy distribution gives the most random solution, *i.e.*, it introduces the minimum additional information beyond what is implied in the original available mean constraints.

It should be pointed out that information theoretic analysis neither pretends to replace the “classical” queueing solutions nor to be an approximation to that “classical” results. The idea is just to apply the maximum entropy formalism in order to get the widest probability distribution subject to the known mean value constraints. Hence, when along the paper we present “classical” queueing results (given in terms of exact analytical formulas or numerical inversions) *versus* maximum entropy solutions, we only wish to display two alternative tools for analyzing an unique real underlying queueing phenomenon. It is so far of our intention to suggest a possible (philosophical or numerical) superiority of the “classical” methodology over the maximum entropy approach or *vice versa*.

Although the general theory (*i.e.*, mathematical properties, numerical techniques) is common for both the discrete case (mass probability functions) and the continuous case (density functions), the numerical effort to carry out the latter is essentially superior. One standard method for maximizing the entropy is to introduce Lagrangian multipliers and then employ numerical methods that reduce the problem to find the solution of a non-linear system of equations or, alternatively, to minimize a potential function (see [8] and the references therein). The consideration of the Newton-Raphson method or the use of direct search algorithms [5] provide the basis for a computer program that obtains the optimal Lagrangian

multipliers and gives the maximum entropy solution given arbitrary expected value constraints.

The numerical implementation in a discrete case (*i.e.*, when the mass probability function takes values on  $\{0, \dots, n\}$  or on  $\mathbb{Z}_+$ ) is a simple task which can be performed in a personal computer consuming a few seconds when running. In contrast, the continuous case implies to estimate a density function taking values over the domain  $(0, +\infty)$ . It typically demands several hours of running time, and so often the program stops without reaching the desired convergence to the optimal Lagrange multipliers. In practice, to solve the entropy maximization in a continuous framework one must make several trials (by using a variety of algorithms and different initial guess multipliers) until arriving to the optimal multipliers.

These demanding computational requirements explain that almost all the existing studies only deal with discrete cases. As an exception, we mention de paper by Wagner and Geyer [24], where the accuracy of the maximum entropy approach is investigated for Gamma, lognormal and other continuous distributions. The application of information theoretic methods to queueing problems is more involved because the continuous measures under study (*i.e.*, a busy period, a waiting time, etc.) are determined by the interaction of several random mechanisms (arrival flow, service times) and the rules governing the queueing model (service discipline, number of channels, priorities, etc.). Hence, applications to queueing imply a major degree of complexity in comparison with the work by Wagner and Geyer who only pretend to estimate a prefixed probability density function belonging to a certain family of probability distributions.

The main purpose of this paper is to contribute to the maximum entropy analysis of the continuous performance characteristics of queueing systems. To this end, we will apply the PME to obtain the density function of the busy period in controllable  $M/G/1$  queueing models. The control of the service process of  $M/G/1$  queues has been extensively studied in the literature. The surveys of Doshi [6] and Teghem [23], and the monograph of Takagi [22] present a complete methodological overview, describe applications of various vacation models and gave appropriate bibliography on the topic. We will restrict our attention to the so-called exhaustive service case (*i.e.*, the system must be empty when the server starts a vacation period). More specifically, we will obtain the maximum entropy density for the busy period of an  $M/G/1$  queue operating under the  $N$ -,  $T$ - and  $D$ -policies. The busy period analysis [9] is an important feature from the server's point of view and it is also helpful in optimization problems leading to an efficient planning of the system resources.

Gelenbe and Iasnogorodski [12] gave the first proof of a decomposition result for a single server queue with general interarrival times and general service and vacation times. The stochastic decomposition is an useful property which permits the system to be analyzed by considering the waiting time for the  $GI/G/1$  system without vacations plus the additional contribution due to the vacations. In the context of our paper we reduce to the busy period analysis; however, since the waiting time is another significant performance measure of continuous type, it could be the subject matter of any subsequent study.

As a second open direction, we mention the application of the PME to continuous characteristics of queueing networks. An interesting possibility could be the investigation of the waiting time in G-networks operating under the presence of an extra flow of negative customers, which imply the delation of work from the system at the negative arrival epochs [2, 10, 11].

The rest of the paper is organized as follows. In Section 2, we summarize mathematics of the maximum entropy formalism. In Section 3 we describe the  $M/G/1$  queue operating under the three controllable policies. For each of the three policies, in Section 4 we obtain the ME density for the busy period given that the available information consists in some mean value constraints (such as the first moments or the value of the Laplace transform at a given point). A brief summary for the “classical” busy period analysis of these vacation models is also given. Numerical examples provide insight on the maximum entropy solutions and the numerical inversion of “classical” results for several traffic intensity levels and different choices of the system parameters  $N$ ,  $T$  and  $D$ . Finally, some conclusions are given in Section 5.

## 2. MATHEMATICS OF THE PRINCIPLE OF MAXIMUM ENTROPY

In this section we summarize some mathematical questions relating to the calculation of ME-density functions. In general, it is difficult or impossible to obtain closed-form solutions. We therefore discuss numerical methods for computing the Lagrangian multipliers.

Let  $f(x)$  be a density function associated to the queueing performance descriptor under study. The density  $f(x)$  takes values in a continuous state space  $\mathcal{X}$ . Thus, we have the normalization condition

$$\int_{\mathcal{X}} f(x) dx = 1. \quad (1)$$

We assume that the information about  $f(x)$  takes the form of linear equality constraints

$$\int_{\mathcal{X}} F_k(x) f(x) dx = F_k, \quad 1 \leq k \leq m, \quad (2)$$

for known functions  $F_k(\cdot)$  and known numbers  $F_k$ . We note that the structural form of the mean value constraints (2) covers important special cases such as: (i)  $F_k(x) = x^k$  (central moments of order  $k$ ); (ii)  $F_k(x) = I_{(-\infty, x_k]}(x)$  (value of the distribution function at the point  $x_k$ ); (iii)  $F_k(x) = e^{-s_k x}$  (value of the Laplace transform at the point  $s_k$ ).

The PME states that, of all the probability density functions satisfying the mean value constraints (1)–(2), the minimal prejudiced density (*i.e.*, the density function that introduces the minimum extraneous information beyond what is implied in the original problem) is the one that maximizes the Shannon’s entropy

functional

$$H(f) = - \int_{\mathcal{X}} f(x) \ln f(x) dx. \quad (3)$$

Suppose that a prior density  $g(x)$  is given as current estimate, then the principle of minimum cross-entropy generalizes the PME by stating that, of all the densities that satisfy the mean constraints, the minimum cross-entropy solution is chosen by minimizing the functional

$$H(f, g) = \int_{\mathcal{X}} f(x) \ln \frac{f(x)}{g(x)} dx. \quad (4)$$

In fact, the PME corresponds to the particular case when the prior density  $g(x)$  in (4) is uniformly distributed on the state space  $\mathcal{X}$ .

The maximization of  $H(f)$  can be carried out with the help of the method of Lagrange's multipliers. If there exists a density function that minimizes the entropy (3) and satisfies the mean value constraints (1)-(2), then it has the following form

$$\hat{f}(x) = \exp \left\{ -\alpha_0 - \sum_{k=1}^m F_k(x) \alpha_k \right\}, \quad x \in \mathcal{X}, \quad (5)$$

where  $\alpha_k$  are the Lagrangian multipliers.  $\alpha_0$  is determined from the normalization condition (1), so we obtain

$$\exp \{ \alpha_0 \} = \int_{\mathcal{X}} \exp \left\{ - \sum_{k=1}^m F_k(x) \alpha_k \right\} dx. \quad (6)$$

The rest of Lagrangian multipliers satisfy the following relations

$$-\frac{\partial \alpha_0}{\partial \alpha_k} = F_k, \quad 1 \leq k \leq m. \quad (7)$$

One can sometimes determine  $\alpha_k$  from the above equation (7). However, in general, it is impossible to solve equations (7) for  $\alpha_k$  explicitly.

The special case where  $m = 1$ ,  $\mathcal{X} = (0, +\infty)$  and  $F_1(x) = x$  yields the explicit ME density

$$f(x) = \frac{1}{F_1} e^{-\frac{x}{F_1}}, \quad x \in \mathcal{X}. \quad (8)$$

Suppose that we add the second moment as a new constraint, then it is impossible to get a closed-form explicit solution for the pair  $(\alpha_1, \alpha_2)$ . Therefore, numerical methods of solution become important. By combining (2) and (5) we observe that standard method for finding the optimal  $\alpha_k$  is to solve the following system [8]:

$$F_i(\alpha_1, \dots, \alpha_m) = \int_{\mathcal{X}} (F_i(x) - F_i) \exp \left\{ - \sum_{k=1}^m (F_k(x) - F_k) \alpha_k \right\} dx = 0, \quad 1 \leq i \leq m. \quad (9)$$

The above equations (9) for the Lagrange multipliers are implicit and non-linear. To solve the system (9) is equivalent to minimizing the potential function

$$F(\alpha_1, \dots, \alpha_m) = \log \int_{\mathcal{X}} \exp \left\{ - \sum_{k=1}^m (F_k(x) - F_k) \alpha_k \right\} dx, \quad (10)$$

or, alternatively, the balanced function

$$G(\alpha_1, \dots, \alpha_m) = \sum_{i=1}^m p_i \left( \int_{\mathcal{X}} (F_i(x) - F_i) \exp \left\{ - \sum_{k=1}^m (F_k(x) - F_k) \alpha_k \right\} dx \right)^2, \quad (11)$$

where  $0 < p_i < 1$  and  $\sum_{i=1}^m p_i = 1$ .

The balanced equation  $G(\alpha_1, \dots, \alpha_m)$  in (11) takes the value 0 at the optimal solution  $(\alpha_1^*, \dots, \alpha_m^*)$  which provides a computational advantage over the potential function (10). For computing the minimum in (11) we will employ Nelder and Mead's algorithm [18] which is a method of direct search and does not involve derivatives, avoiding the problems arising when the Hessian of  $G(\alpha_1, \dots, \alpha_m)$  is algorithmically almost singular. The method remains valid for any arbitrary number of constraints but, the search mechanism requires an increasingly number of evaluations on the function  $G$  and, obviously the algorithmic convergence to the optimal multipliers may be really slow.

As a practical remark, we mention that the numerical computation of the integrals involved in (11) implies the consideration of a finite truncated domain (usually of the form  $(0, b)$  where the upper limit  $b$  is determined with the help of Tchebychev's inequality) and the use of the NRS (Numerical Recipes Software) subroutine MIDPNT [19].

### 3. MODEL DESCRIPTION OF THE $M/G/1$ VACATION MODELS UNDER STUDY

We consider single server queueing systems to which customers arrive according to a Poisson stream of rate  $\lambda$ . The service times are mutually independent and independent of the arrival process with common probability distribution function  $B(x)$  ( $B(0) = 0$ ),  $k$ -th moment  $\beta_k$ , Laplace transform  $\beta(s)$  and traffic intensity  $\rho = \lambda\beta_1$ . The service facility consists in the queue and the server. The server is turned off as soon as the service facility becomes empty, *i.e.*, we consider only the case of exhaustive vacations. We assume that  $\rho < 1$  which implies that the system is in the steady state.

We will consider the three following classical policies for switching on the server:

- (a) *N-policy* [26]: the server is turned on when the queue length reaches the number  $N$ .
- (b) *T-policy* [14, 17]: the server checks the queue after  $T$  time units. If the server finds a non empty queue then it automatically initiates service;

otherwise, the server takes a second vacation period of length  $T$ . This mechanism is repeated until the first time in which the server finds at least one customer in the queue.

- (c) *D-policy* [4]: the server is turned on when the cumulative service time of all arrived customers reaches a prefixed threshold  $D$ .

In what follows we will restrict ourselves to the above controllable policies. However, many variants and generalizations could be considered. For example, after a vacation of length  $T$  finding an empty system one possibility is to take another identical vacation but, a second option is to wait for the first arrival and then to turn on the server. By combining the three simple policies we obtain dyadic policies such as the  $Min(N, T)$ -policy or the  $Max(N, D)$ -policy.

The “classical” analysis of the three aforementioned policies forms an integral part of the queueing literature. Recently, Wang *et al.* [25] use the PME to study the  $M/G/1$  queue with  $N$ -policy. However, an analysis based on information theoretic techniques to obtain specific continuous performance characteristics of queueing models is almost inexistent. The aim of this paper is to fill this gap by applying the PME to the busy period analysis of the  $M/G/1$  queue operating under the  $N$ -,  $T$ - and  $D$ -policies.

To conclude this section we introduce some notation. In the context of queueing systems with server vacations, a busy period  $L$  is defined as the length of time elapsed from the epoch when the server returns to the system to provide service until the first departure epoch leaving an empty system behind. Let  $L_C$ ,  $L_N$ ,  $L_T$  and  $L_D$  be random variables denoting the busy period of the “classical”  $M/G/1$  queue and the controllable  $M/G/1$  queue with  $N$ -,  $T$ - and  $D$ - policies, respectively.

## 4. CLASSICAL AND MAXIMUM ENTROPY SOLUTIONS FOR THE BUSY PERIOD OF CONTROLLABLE $M/G/1$ QUEUES

### 4.1. THE $M/G/1$ QUEUE OPERATING UNDER THE $N$ -POLICY

In this policy introduced by Yadin and Naor [26] the server returns to service when there are  $N$  customers present for the first time after its removal. Thus,  $L_N$  is the sum of  $N$  independent and identically distributed busy periods of a standard  $M/G/1$  queueing system. Hence, its Laplace transform  $\beta_{L_N}(s)$  is given by

$$\beta_{L_N}(s) = (\beta_{L_C}(s))^N, \quad (12)$$

where  $\beta_{L_C}(s)$  is the Laplace transform of the standard queue, *i.e.*,  $\beta_{L_C}(s) = \beta(s + \lambda - \lambda\beta_{L_C}(s))$ .

From (12) and the expressions for the moments of the busy period in the standard  $M/G/1$  queue [15], we find that

$$E[L_N] = \frac{N\beta_1}{1-\rho}, \quad (13)$$

$$E[L_N^2] = \frac{N(N-1)\beta_1^2}{(1-\rho)^2} + \frac{N\beta_2}{(1-\rho)^3}, \quad (14)$$

$$E[L_N^3] = \frac{N(N-1)(N-2)\beta_1^3}{(1-\rho)^3} + \frac{N(3(N-1)\beta_1\beta_2 + \beta_3)}{(1-\rho)^4} + \frac{3N\lambda\beta_2^2}{(1-\rho)^5}. \quad (15)$$

The “classical” approach does not provide an explicit expression for the density of the busy period. In the particular case where the service times are exponentially distributed with rate  $\nu$ , the expression (12) can be inverted and the corresponding density function has the following form

$$f_{L_N}(x) = \frac{e^{-(\lambda+\nu)x} N I_N(2x\sqrt{\lambda\nu})}{x\rho^{N/2}}, \quad x > 0, \quad (16)$$

where  $I_N(\cdot)$  denotes the modified Bessel function of order  $N$ , *i.e.*,  $I_N(z) = \sum_{i=0}^{\infty} \frac{1}{i!(i+N)!} \left(\frac{z}{2}\right)^{N+2i}$ .

We next illustrate numerically the use of the PME to estimate the density function of the busy period in  $M/G/1$  queues operating under the  $N$ -policy. Firstly, we will assume that the available information consists in the first and second moments of  $L_N$ . For simplicity, we assume that such moments are given by the “classical” expressions (13) and (14). Alternatively, the use of empirical data to construct constraints based on sample moments provides another choice consistent with the philosophical principles of the maximum entropy approach. Once the mean constraints are fixed, the entropy formalism is independent of the service time distribution. Thus, the choice of  $B(x)$  is not an important matter here.

We now consider service times hyperexponentially distributed having mean value  $\beta_1 = 0.25$  and coefficient of variation  $C_B = (\beta_2 - \beta_1^2)^{1/2} / \beta_1 = 1.25$ . Then, we choose the arrival rate to fix the traffic at levels  $\rho = 0.25$  and  $0.75$ . The number of customers required to turn on the server is fixed at levels  $N = 2, 10$  and  $50$ . By combining  $\rho$  and  $N$ , we have six different scenarios. In all cases, the ME density  $\hat{f}_{L_N}^2(x)$  is numerically computed with the help of the methodology described in Section 2. Note that the ME density satisfies the given constraints, so the first two moments of  $\hat{f}_{L_N}^2(x)$  and  $f_{L_N}(x)$  are coincident. Thus, we next compare the “classical” and the ME solutions with the help of the following two measures:

$$\mathcal{D}_3^N = \left| 1 - \frac{E[\hat{L}_N^3]}{E[L_N^3]} \right|, \quad (17)$$

$$\mathcal{D}^N(s) = \left| 1 - \frac{\beta_{\hat{L}_N}(s)}{\beta_{L_N}(s)} \right|. \quad (18)$$



TABLE 1.  $M/H_2/1$  queues with  $N$ -policy: comparing the 3rd moment.

		$N = 2$	$N = 10$	$N = 50$
$\rho = 0.25$	$E[L_N^3]$	2.594	70.007	5331.520
	$E[\widehat{L}_N^3]$	2.453	65.164	5300.938
	$\mathcal{D}_3^N$	0.054	0.069	0.005
$\rho = 0.75$	$E[L_N^3]$	630.374	6331.874	207159.3
	$E[\widehat{L}_N^3]$	745.301	5366.929	196052.5
	$\mathcal{D}_3^N$	0.182	0.152	0.053

TABLE 2. Comparing the Laplace transforms: the case  $M/H_2/1$  with  $N = 10$ .

$N = 10$						
$\rho$	0.25			0.75		
$s$	$\beta_{L_N}(s)$	$\beta_{\widehat{L}_N}(s)$	$\mathcal{D}^N(s)$	$\beta_{L_N}(s)$	$\beta_{\widehat{L}_N}(s)$	$\mathcal{D}^N(s)$
0.001	0.996673	0.996673	0.000000	0.990095	0.990095	0.000000
0.01	0.967344	0.967345	0.000000	0.908684	0.908804	0.000132
0.05	0.849224	0.849304	0.000094	0.657855	0.663731	0.008932
0.1	0.725539	0.726056	0.000713	0.475406	0.494942	0.041093
0.3	0.406262	0.412535	0.015440	0.185914	0.244190	0.313459
0.5	0.242017	0.256333	0.059152	0.093260	0.161867	0.735655
0.7	0.151015	0.171883	0.138183	0.052686	0.121026	1.29709
0.9	0.097726	0.122779	0.256361	0.032040	0.096635	2.01607
1.2	0.053692	0.081393	0.515903	0.016665	0.074199	3.45217
1.5	0.031057	0.058719	0.890670	0.009370	0.060216	5.42650
2.0	0.013615	0.038725	1.84419	0.004041	0.045823	10.3391

Formula (17) for  $\mathcal{D}_3^N$  gives idea of a relative distance between the third order moments of the maximum entropy and the “classical” distributions. The quantity  $\mathcal{D}^N(s)$  in (18) compares both solutions in terms of the Laplace transforms at a given point  $s$ .

Table 1 shows that  $E[L_N^3]$  and  $E[\widehat{L}_N^3]$  are increasing functions of  $N$  and  $\rho$ . For each fixed value of  $N$ , the relative distance  $\mathcal{D}_3^N$  increases as function of the traffic intensity  $\rho$ .

In Table 2 we compare the “classical” and ME densities for the case  $N = 10$ . To this end, we vary the auxiliary parameter  $s$  and compute  $\beta_{L_N}(s)$ ,  $\beta_{\widehat{L}_N}(s)$  and  $\mathcal{D}^N(s)$ . It can be observed that  $\mathcal{D}^N(s)$  is an increasing function of  $s$  and  $\rho$ , *i.e.*, “classical” and ME solutions are closer when the traffic intensity decreases. The increment of  $\mathcal{D}^N(s)$  for increasing values of  $s$  is an expected result because the ME density  $\widehat{f}_{L_N}^2(x)$  is based only on the constraints  $E[L_N]$  and  $E[L_N^2]$ , which

TABLE 3.  $M/M/1$  queues with  $N$ -policy: skewness, shape, absolute distance.

	$N = 2$		$N = 10$		$N = 50$	
$\rho = 0.25$	2.40	<b>1.55</b>	1.08	<b>0.13</b>	0.48	$-4.7 \times 10^{-5}$
	BS	<b>D</b>	BS	<b>BS</b>	BS	<b>BS</b>
	0.218060		0.262078		0.121546	
$\rho = 0.5$	3.58	<b>4.84</b>	1.58	<b>0.41</b>	0.71	$1.5 \times 10^{-3}$
	BS	<b>D</b>	BS	<b>BS</b>	BS	<b>BS</b>
	0.184811		0.348290		0.179111	
$\rho = 0.75$	5.57	<b>7.05</b>	2.49	<b>1.26</b>	1.11	$8.2 \times 10^{-2}$
	BS	<b>D</b>	BS	<b>D</b>	BS	<b>BS</b>
	0.329070		0.363570		0.279231	

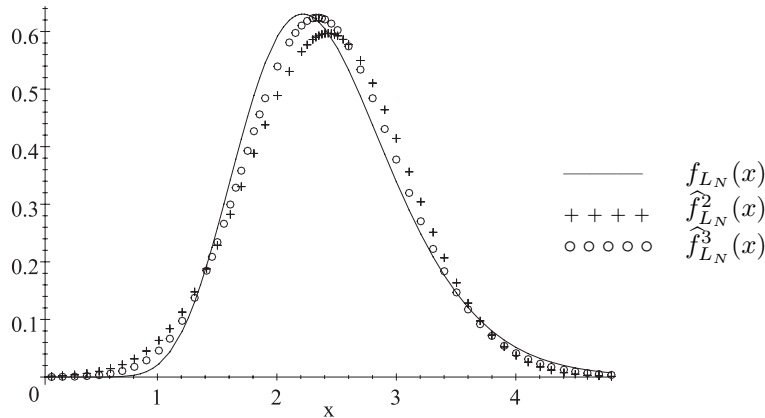


FIGURE 1.  $M/M/1$  queue with  $N = 17$ : “classical” and ME densities.

are determined by the derivatives of the Laplace transform of  $L_N$  at the origin  $s = 0$ . Later on we will show how to improve the maximum entropy estimation by adding more constraints.

In a second set of numerical results (see Tab. 3 and Fig. 1) we turn our attention to the pure Markovian queue with  $N$ -policy. We reduce to this special case to have the explicit expression (16) for the “classical” density  $f_{L_N}(x)$ . In Table 3, we consider that  $\beta_1 = 0.25$  and  $\lambda$  varies to fix the traffic intensity at values  $\rho = 0.25, 0.5$  and  $0.75$ . Each cell in the table contains five entries. The two numbers in the first row of the cell correspond to the skewness coefficient defined by  $E[(L_N - E[L_N])^3] / (\text{Var}(L_N))^{3/2}$  associated to the “classical” density  $f_{L_N}(x)$  and the ME density  $\hat{f}_{L_N}^2(x)$  (written in bold). The second row indicates the shape of the density. The notations D and BS mean “decreasing” and “bell-shape”, respectively. Bold characters again correspond to the ME density. The number in the last row of each cell is the global absolute distance between both densities

defined by

$$D(f_{L_N}, \widehat{f}_{L_N}^2) = \int_0^\infty |f_{L_N}(x) - \widehat{f}_{L_N}^2(x)| dx. \tag{19}$$

This quantity provides another useful global measure of proximity between  $f_{L_N}(x)$  and  $\widehat{f}_{L_N}^2(x)$ . We note that the asymmetry of both densities decreases for large values of  $N$  and increases with  $\rho$ . Looking at the queues with bell-shape ME densities, we observe that queues with lesser skewness coefficients yield to smaller absolute distances.

Up to this point we have used the third moment only to give a measure of proximity between the “classical” and the ME distributions (see  $\mathcal{D}_3^N$  in Tab. 1). Next we show how the use of  $E[L_N^3]$  as an extra constraint leads to ME densities  $\widehat{f}_{L_N}^3(x)$  closer to the “classical” density  $f_{L_N}(x)$ . To this end, we now consider an  $M/M/1$  queue with traffic intensity  $\rho = 0.125$  and  $N = 17$ . Figure 1 displays the “classical” density  $f_{L_N}(x)$  and the ME solutions  $\widehat{f}_{L_N}^2(x)$  and  $\widehat{f}_{L_N}^3(x)$  based on the first two and three moments, respectively. The global absolute distances  $D(f_{L_N}, \widehat{f}_{L_N}^2) = 0.1663$  and  $D(f_{L_N}, \widehat{f}_{L_N}^3) = 0.0934$  quantify that  $\widehat{f}_{L_N}^3(x)$  is considerably closer to  $f_{L_N}(x)$  than  $\widehat{f}_{L_N}^2(x)$ .

We now come back to the  $M/H_2/1$  queue with parameters  $\beta_1 = 0.25$ ,  $C_B = 1.25$  and  $N = 10$ . The objective is to improve the ME solution but this time we consider that the new constraint is the value of the Laplace transform at any point  $s_0$ . In this way, we combine information about  $\beta_{L_N}(s)$  at the origin  $s = 0$  (given by the first two moments) and information at another different point  $s_0$ . The ME formalism described in Section 2 can be adapted leading to the ME density

$$\widehat{f}_{L_N}^{2,1}(x) = \exp\{-\alpha_0 - \alpha_1 x - \alpha_2 x^2 - \gamma e^{-s_0 x}\}, \quad x > 0. \tag{20}$$

An explicit expression for the “classical” density  $f_{L_N}(x)$  of the model  $M/H_2/1$  with  $N$ -policy is unknown. However, we can carry out the numerical inversion of  $\beta_{L_N}(s)$  in (12) using the algorithm POST-WIDDER described in Abate and Whitt [1]. In Figure 2, we consider the case  $\rho = 0.25$  and display three curves which correspond to the numerical inversion denoted by  $\widetilde{f}_{L_N}(x)$  and the ME densities  $\widehat{f}_{L_N}^2(x)$  and  $\widehat{f}_{L_N}^{2,1}(x)$ . The latter was obtained by taking as extra constraint the value of  $\beta_{L_N}(s)$  at  $s_0 = 0.7$ . The curves indicate that  $\widehat{f}_{L_N}^{2,1}(x)$  is considerable closer to  $\widetilde{f}_{L_N}(x)$  than  $\widehat{f}_{L_N}^2(x)$ . The same conclusion is obtained in Figure 3, for a traffic intensity  $\rho = 0.75$  where the point  $s_0$  is chosen as 1.5. In fact, in this example we observe that  $\widehat{f}_{L_N}^2(x)$  is a decreasing function but this shape is changed when we add the constraint  $\beta_{L_N}(s_0)$ , in agreement with the shape of the “classical” inverted density  $\widetilde{f}_{L_N}(x)$ .

#### 4.2. THE $M/G/1$ QUEUE OPERATING UNDER THE $T$ -POLICY

The  $T$ -policy [14, 17] consists of turning the server on after  $T$  time units have elapsed from the epoch of server removal, if there is at least one customer present

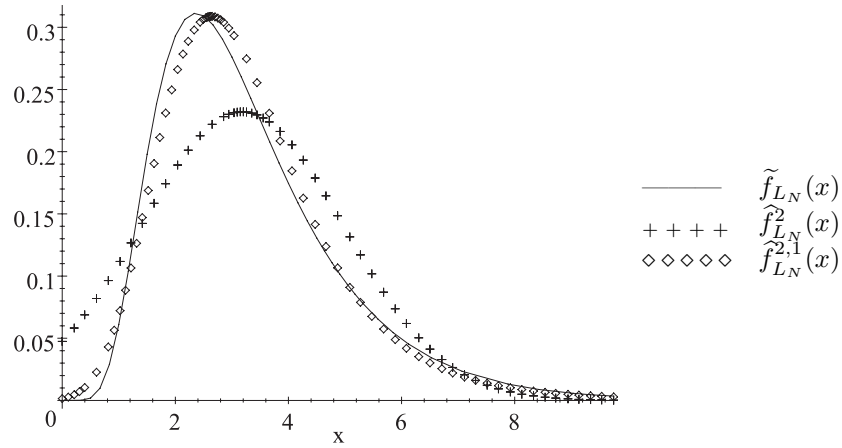


FIGURE 2.  $M/H_2/1$  queue with  $N = 10$ ,  $\rho = 0.25$ : “classical” and ME densities.

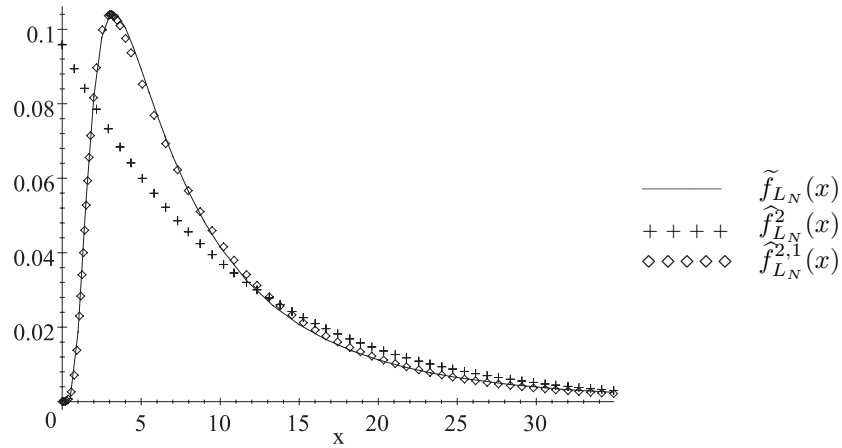


FIGURE 3.  $M/H_2/1$  queue with  $N = 10$ ,  $\rho = 0.75$ : “classical” and ME densities.

in the waiting line. The server remains on until the queue becomes empty again. If after  $T$  time units there are no customers in the system to initiate service, the server takes another  $T$  time units vacation period, and so on, until at least one customer is present.

The relation between the Laplace transforms of  $L_T$  and  $L_C$  is

$$\beta_{L_T}(s) = \frac{e^{-\lambda T} (e^{\lambda T \beta_{L_C}(s)} - 1)}{1 - e^{-\lambda T}}. \tag{21}$$

TABLE 4.  $M/H_2/1$  queues with  $T$ -policy: comparing the 3rd moment.

		$T = 2.5$	$T = 5.0$	$T = 25.0$
$\rho = 0.25$	$E[L_T^3]$	5.714	18.834	838.908
	$E[\widehat{L}_T^3]$	5.805	17.394	818.219
	$\mathcal{D}_3^T$	0.016	0.076	0.024
$\rho = 0.75$	$E[L_T^3]$	4287.214	14560.32	616176.5
	$E[\widehat{L}_T^3]$	4253.440	12644.20	596482.5
	$\mathcal{D}_3^T$	0.007	0.131	0.031

Differentiating (21), we obtain

$$E[L_T] = \frac{T\rho}{(1-\rho)(1-e^{-\lambda T})}, \tag{22}$$

$$E[L_T^2] = \frac{1}{1-e^{-\lambda T}} \left( \frac{T^2\rho^2}{(1-\rho)^2} + \frac{T\lambda\beta_2}{(1-\rho)^3} \right), \tag{23}$$

$$E[L_T^3] = \frac{\lambda T}{1-e^{-\lambda T}} \left( \frac{T^2\rho^2\beta_1}{(1-\rho)^3} + \frac{3T\rho\beta_2 + \beta_3}{(1-\rho)^4} + \frac{3\lambda\beta_2^2}{(1-\rho)^5} \right). \tag{24}$$

Again a “classical” closed-form formula for the density of the busy period  $L_T$  does not exist. In the particular case of exponential service times the expression (21) can be inverted and the density of  $L_T$  has the form (minor correction of formula (50) in [9]):

$$f_{L_T}(x) = \frac{e^{-(\lambda+\nu)x} e^{-\lambda T} T \sqrt{\lambda\nu} I_1 \left( 2\sqrt{\lambda\nu x(x+T)} \right)}{(1-e^{-\lambda T}) \sqrt{x(x+T)}}, \quad x > 0. \tag{25}$$

In Tables 4 and 5 we deal with the  $M/H_2/1$  queue with  $T$ -policy. We again assume that  $\beta_1 = 0.25$  and  $C_B = 1.25$ . The parameter  $T$  takes values 2.5, 5.0 and 25.0. In Table 4 we present  $E[L_T^3]$ ,  $E[\widehat{L}_T^3]$  and  $\mathcal{D}_3^T$ . Similar comments to those presented for  $N$ -policy remain valid here. Laplace transforms given in Table 5 for the case  $T = 5.0$  show that measure  $\mathcal{D}^T(s)$  is smaller when the traffic intensity decreases.

Following a parallel study to that performed for the  $N$ -policy, in Table 6 and Figure 4 we consider the case of exponential service times. In Table 6, we assume that  $\beta_1 = 0.25$  and arrival intensities are fixed to obtain  $\rho = 0.25, 0.5$  and  $0.75$ . The table provides information about the skewness and the shape of  $f_{L_T}(x)$  (given by expression (25)) and  $\widehat{f}_{L_T}^2(x)$ . Similarly to formula (19), we define the global absolute distance  $D(f_{L_T}, \widehat{f}_{L_T}^2)$  which is given in the last entry of each cell.

Figure 4 shows the gain obtained by using the third moment  $E[L_T^3]$  (see (24)) as a known constraint. To this end, we choose  $\lambda = 1.0, \nu = 8.0, T = 15.0$  and display the “classical” density  $f_{L_T}(x)$  versus  $\widehat{f}_{L_T}^2(x)$  and  $\widehat{f}_{L_T}^3(x)$ . The improvement can

TABLE 5. Comparing the Laplace transforms: the case  $M/H_2/1$  with  $T = 5.0$ .

$T = 5.0$						
$\rho$	0.25			0.75		
$s$	$\beta_{L_T}(s)$	$\beta_{\hat{L}_T}(s)$	$\mathcal{D}^T(s)$	$\beta_{L_T}(s)$	$\beta_{\hat{L}_T}(s)$	$\mathcal{D}^T(s)$
0.001	0.998324	0.998324	0.000000	0.985187	0.985187	0.000000
0.01	0.983453	0.983453	0.000000	0.866796	0.867024	0.000263
0.05	0.921618	0.921643	0.000027	0.540541	0.549978	0.017458
0.1	0.852967	0.853134	0.000196	0.341325	0.368070	0.078356
0.3	0.647687	0.650080	0.003696	0.097996	0.153057	0.561869
0.5	0.512872	0.519267	0.012469	0.042090	0.095855	1.27738
0.7	0.418728	0.429541	0.025822	0.021826	0.069690	2.19298
0.9	0.349965	0.364877	0.042609	0.012690	0.054727	3.31261
1.2	0.276568	0.296503	0.072080	0.006476	0.041386	5.39067
1.5	0.225487	0.249060	0.104589	0.003709	0.033271	7.97034
2.0	0.168812	0.196082	0.161540	0.001737	0.025074	13.4352

TABLE 6.  $M/M/1$  queues with  $T$ -policy: skewness, shape, absolute distance.

	$T = 2.5$		$T = 5.0$		$T = 25.0$	
$\rho = 0.25$	1.93	<b>1.72</b>	1.38	<b>0.89</b>	0.61	<b><math>3.3 \times 10^{-2}</math></b>
	BS	<b>D</b>	BS	<b>BS</b>	BS	<b>BS</b>
	0.044726		0.150433		0.154493	
$\rho = 0.5$	2.01	<b>1.46</b>	1.42	<b>0.63</b>	0.63	<b><math>9.0 \times 10^{-3}</math></b>
	BS	<b>D</b>	BS	<b>BS</b>	BS	<b>BS</b>
	0.145215		0.247946		0.161318	
$\rho = 0.75$	2.71	<b>2.17</b>	1.91	<b>0.91</b>	0.85	<b><math>3.3 \times 10^{-2}</math></b>
	BS	<b>D</b>	BS	<b>BS</b>	BS	<b>BS</b>
	0.163159		0.303207		0.216466	

be measured in terms of the global absolute distances  $D(f_{L_T}, \hat{f}_{L_T}^2) = 0.1596$  and  $D(f_{L_T}, \hat{f}_{L_T}^3) = 0.0944$ .

Finally, in Figures 5 and 6, we improve the maximum entropy estimation of  $L_T$  by using as a given information the value of  $\beta_{L_T}(s)$  (see formula (21)) at the point  $s_0$ . Figures 5 and 6 deal with the  $M/H_2/1$  queue with  $T = 5.0$  and  $\rho = 0.25$  and  $0.75$ , respectively. In the case  $\rho = 0.75$  we illustrate the dependence of  $\hat{f}_{L_T}^{2,1}(x)$  on the point  $s_0$ . To this end, we display  $\tilde{f}_{L_T}(x)$ ,  $\hat{f}_{L_T}^2(x)$  and three more curves corresponding to  $\hat{f}_{L_T}^{2,1}(x)$  for different choices of  $s_0$ . In Figure 5,  $s_0$  has been chosen as 1.2.

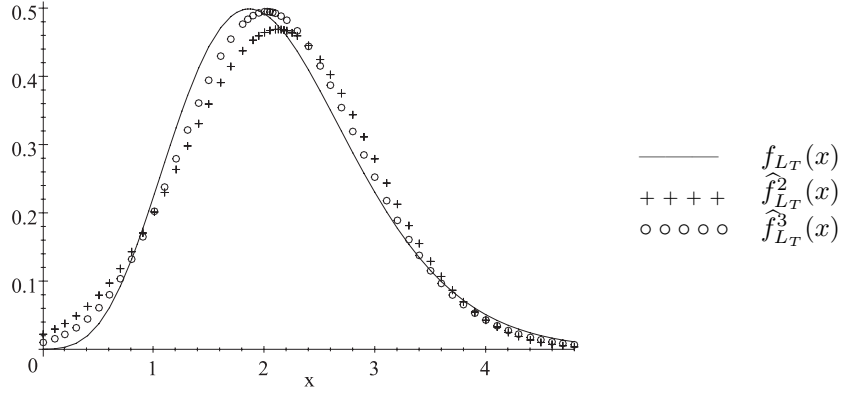


FIGURE 4.  $M/M/1$  queue with  $T = 15$ : “classical” and ME densities.

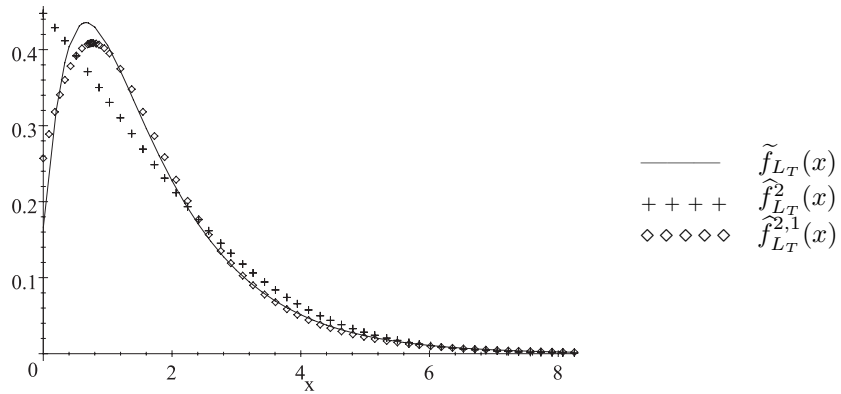


FIGURE 5.  $M/H_2/1$  queue with  $T = 5$ ,  $\rho = 0.25$ : “classical” and ME densities.

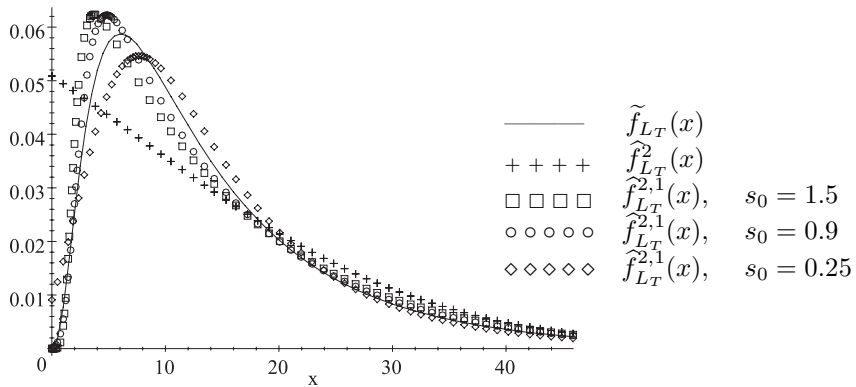


FIGURE 6.  $M/H_2/1$  queue with  $T = 5$ ,  $\rho = 0.75$ : “classical” and ME densities.

4.3. THE  $M/G/1$  QUEUE OPERATING UNDER THE  $D$ -POLICY

The  $D$ -policy is a more sophisticated policy that takes into account the service times of the waiting customers to switch on the server. To this end, the server is able to know each customer's workload. The removed server is turned on as soon as the total workload for all customers in the waiting line firstly exceeds the threshold  $D$ .

Following formula (6.5) in Artalejo [3] we observe that

$$\beta_{L_T}(s) = \beta_Z(s + \lambda - \lambda\beta_{L_C}(s)), \quad (26)$$

where  $\beta_Z(s)$  denotes the Laplace transform of the total work accumulated just at the epoch when the server is switched on. The corresponding distribution function is given by

$$F_Z(x) = B(x) - \int_0^D (1 - B(x - y)) dm(y), \quad x > D, \quad (27)$$

and  $m(x) = \sum_{k=1}^{\infty} B^k(x)$  is the renewal function associated with the renewal process formed by a sequence of service times, so  $B^k(x)$  is the  $k$ -fold self-convolution of  $B(x)$ .

The two first moments of  $L_D$  are given by

$$E[L_D] = \frac{(1 + m(D))\beta_1}{1 - \rho}, \quad (28)$$

$$E[L_D^2] = \frac{(1 + m(D))\beta_2}{(1 - \rho)^3} + \frac{2\beta_1}{(1 - \rho)^2} \int_0^D x dm(x). \quad (29)$$

Gakis *et al.* give a complicate expression (see formula (28) in [9]) for the density function of  $L_D$  in terms of convolutions of the density of a busy period in the standard  $M/G/1$  queue. The next expression for the case of exponential service times corrects a minor mistake in formula (53) of [9]:

$$f_{L_D}(x) = \nu e^{\nu D} e^{-(\lambda + \nu)x} \left( 1 + \int_D^x \frac{\lambda y}{x - y} \sqrt{\frac{x - y}{\rho x}} I_1 \left( 2\sqrt{\lambda \nu x(x - y)} \right) dy \right), \quad x > D. \quad (30)$$

In the exponential case  $m(x) = \nu x$ . Thus, expressions (28) and (29) reduce to the following

$$E[L_D] = \frac{1 + \nu D}{\nu(1 - \rho)}, \quad (31)$$

$$E[L_D^2] = \frac{2(1 + \nu D)}{\nu^2(1 - \rho)^3} + \frac{D^2}{(1 - \rho)^2}. \quad (32)$$

Note that the density function of  $L_D$  takes values on the interval  $(D, +\infty)$ . This remark should be taken into account when we deal with the maximum entropy approach. Accordingly, the integrals arising in the numerical calculation of the Lagrangian multipliers are defined over  $(D, +\infty)$ .



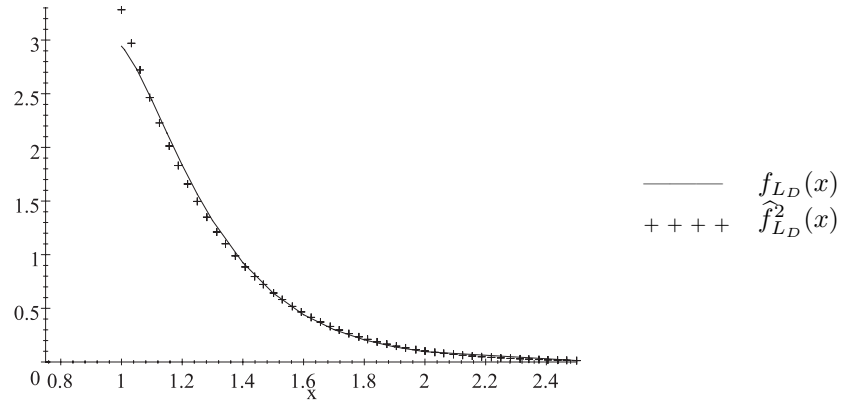


FIGURE 7.  $M/M/1$  queue with  $D = 1$ : “classical” and ME densities.

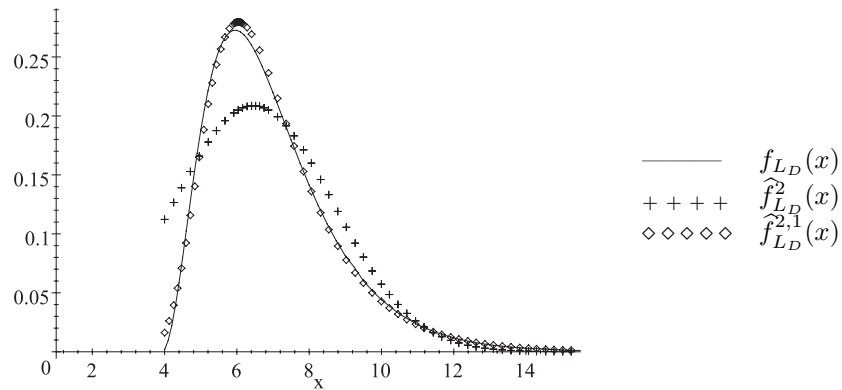


FIGURE 8.  $M/M/1$  queue with  $D = 4$ : “classical” and ME densities.

We now present some numerical results. The “classical” density  $f_{L_D}(x)$  is available in explicit form only for the case of exponential service times (see formula (30)). Furthermore, the numerical inversion of the Laplace transform  $\beta_{L_D}(s)$  has need of knowledge of the renewal function but it is well-known that  $m(x)$  has an explicit expression only for a few particular cases. Due to these reasons, we next restrict to the model  $M/M/1$  with  $D$ -policy.

In Figure 7, we consider a light traffic system with parameters  $\lambda = 1.0$ ,  $\nu = 8.0$  and  $D = 1.0$ . Then, we display the “classical” density  $f_{L_D}(x)$  and the ME density  $\hat{f}_{L_D}^2(x)$  based on the first two moments given by formulas (31) and (32). At the light of the decreasing shape of both densities and the global absolute distance  $D(f_{L_D}, \hat{f}_{L_D}^2) = 0.0482$ , we conclude that both solutions are enough close.

In a second numerical example, we increase  $\rho$  and the threshold  $D$ . Concretely, we consider the parameters  $\lambda = 2.0$ ,  $\nu = 5.0$  and  $D = 4.0$ . Figure 8 shows that it

could be convenient to improve the ME density  $\widehat{f}_{L_D}^2(x)$ . It is done by considering the value of  $\beta_{L_D}(s)$  at the point  $s_0 = 1.0$ . The gain measured in terms of absolute distances yield to the values  $D(f_{L_D}, \widehat{f}_{L_D}^2) = 0.2447$  and  $D(f_{L_D}, \widehat{f}_{L_D}^{2,1}) = 0.0396$ .

## 5. CONCLUSIONS

In this paper we have shown how the maximum entropy formalism provides an elegant method for the investigation of continuous performance measures of queueing models with vacations. Our numerical results show that the use of the first two moments of the busy period and the value of the Laplace transform at a given point is, in general, sufficient to obtain accurate estimations. This study is complementary to the existing literature which in most cases reduces the use of information theoretic methods to the discrete case.

*Acknowledgements.* This research was supported by the project BFM2002-02189.

## REFERENCES

- [1] J. Abate and W. Whitt, Numerical inversion of Laplace transforms of probability distributions. *ORSA J. Comput.* **7** (1995) 36–43.
- [2] J.R. Artalejo, G-networks: A versatile approach for work removal in queueing networks. *Eur. J. Oper. Res.* **126** (2000) 233–249.
- [3] J.R. Artalejo, On the M/G/1 queue with D-policy. *Appl. Math. Modelling* **25** (2001) 1055–1069.
- [4] K.R. Balachandran and H. Tijms, On the D-policy for the M/G/1 queue. *Manage. Sci.* **21** (1975) 1073–1076.
- [5] B.D. Bunday, *Basic Optimization Methods*. Edward Arnold, London (1984).
- [6] B.T. Doshi, Queueing systems with vacations - A survey. *Queue. Syst.* **1** (1986) 29–66.
- [7] M.A. El-Affendi and D.D. Kouvatso, A maximum entropy analysis of the M/G/1 and G/M/1 queueing systems at equilibrium. *Acta Inform.* **19** (1983) 339–355.
- [8] G.I. Falin, M. Martin and J.R. Artalejo, Information theoretic approximations for the M/G/1 retrial queue. *Acta Inform.* **31** (1994) 559–571.
- [9] K.G. Gakis, H.K. Rhee and B.D. Sivazlian, Distributions and first moments of the busy period and idle periods in controllable M/G/1 queueing models with simple and dyadic policies. *Stoch. Anal. Appl.* **13** (1995) 47–81.
- [10] E. Gelenbe, Product-form queueing networks with negative and positive customers. *J. Appl. Prob.* **28** (1991) 656–663.
- [11] E. Gelenbe, G-networks: A unifying model for neural and queueing networks. *Ann. Oper. Res.* **48** (1994) 433–461.
- [12] E. Gelenbe and R. Iasnogorodski, A queue with server of walking type (autonomous service). *Annales de l'Institut Henry Poincaré, Series B* **16** (1980) 63–73.
- [13] S. Guiasu, Maximum entropy condition in queueing theory. *J. Opl. Res. Soc.* **37** (1986) 293–301.
- [14] D. Heyman, The T-policy for the M/G/1 queue. *Manage. Sci.* **23** (1977) 775–778.
- [15] L. Kleinrock, *Queueing Systems, Volume 1: Theory*. John Wiley & Sons, Inc., New York (1975).
- [16] D.D. Kouvatso, Entropy maximisation and queueing networks models. *Ann. Oper. Res.* **48** (1994) 63–126.

- [17] Y. Levy and U. Yechiali, Utilization of idle time in an M/G/1 queueing system. *Manage. Sci.* **22** (1975) 202–211.
- [18] J.A. Nelder and R. Mead, A simplex method for function minimization. *Comput. J.* **7** (1964) 308–313.
- [19] W.H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, *Numerical Recipes in Fortran, The Art of Scientific Computing*. Cambridge University Press (1992).
- [20] J.E. Shore, Information theoretic approximations for M/G/1 and G/G/1 queueing systems. *Acta Inform.* **17** (1982) 43–61.
- [21] L. Tadj and A. Hamdi, Maximum entropy solution to a quorum queueing system. *Math. Comput. Modelling* **34** (2001) 19–27.
- [22] H. Takagi, *Queueing Analysis*. Vol. 1–3, North-Holland, Amsterdam (1991).
- [23] J. Teghem Jr., Control of the service process in a queueing system. *Eur. J. Oper. Res.* **23** (1986) 141–158.
- [24] U. Wagner and A.L.J. Geyer, A maximum entropy method for inverting Laplace transforms of probability density functions. *Biometrika* **82** (1995) 887–892.
- [25] K.H. Wang, S.L. Chuang and W.L. Pearn, Maximum entropy analysis to the N policy M/G/1 queueing systems with a removable server. *Appl. Math. Modelling* **26** (2002) 1151–1162.
- [26] M. Yadin and P. Naor, Queueing systems with a removable server station. *Oper. Res. Quar.* **14** (1963) 393–405.