The Korean Communications
in Statistics Vol. 4, No. 2, 1997
pp. 497 - 505

# Maximum Entropy Principle for Queueing Theory

**SungJin Ahn[1], DongHoon Lim[2] and SooTaek Kim[3]**

## Abstract

We attempt to get a probabilistic model of a queueing system in the maximum entropy condition. Applying the maximum entropy principle to the queueing system, we obtain the most uncertain probability model compatible with the available information expressed by moments.

## 1. Introduction

Queueing theory studies queueing systems by formulating mathematical models of their operation and then using these models to derive measures of performance. The term queueing is used to describe a large class of phenomena involving arrivals, waiting, servicing, and departures. The basic elements of a queueing model depend on, among others, the following factors: arrival distribution, service time distribution, and queue size. Assuming that the queueing system is in a steady-state condition, queueing theory has tended to focus largely on predicting such characteristics of the queue as the waiting time or system state under certain arrivals or service-time distributions.

But generally, we do not know the real probability distributions. The available information is frequently summarized in mean values or higher moments: mean arrival rates, mean service rates, or mean number of customers in the system. Such a situation make us to rely on some other methods to analyse the queueing system. Guiasu [1] proposed the maximum entropy principle as one of such methods. We will modify and extend his propositions.

Suppose we know that a system has a set of possible states $x_k$ with unknown probabilities $f(x_k)$, and we then learn constraints on the distribution $f$: either values of certain

---

1) Professor, Department of Statistics, and Information & Telecommunication Research Center, Gyeongsang National Univ., Chinju 660-701, Korea.
2) Assistant Professor, Department of Statistics, and Information & Telecommunication Research Center, Gyeongsang National Univ., Chinju 660-701, Korea.
3) Professor, Department of Statistics, and Information & Telecommunication Research Center, Gyeongsang National Univ., Chinju 660-701, Korea.

expectations $\sum_k f(x_k) g_i(x_k)$ or bounds on these values. Suppose that we need to choose a distribution $f$ given what we know. Usually there remains an infinite set of distributions that are not ruled out by the constraints. Which one should we choose?

The principle of maximum entropy states that, of all the distributions $f$ that satisfies the constraints, you should choose the one with the largest entropy $-\sum_k f(x_k)\log f(x_k)$. Entropy maximization was first proposed as a general inference procedure by Jaynes [2]. It has been applied successfully in a remarkable variety of fields, as mentioned in [6].

For the maximum entropy principle to be asserted as a general method of inductive inference, it is reasonable to require that different ways of using it to take the same information into account should lead to consistent results. Shore and Johnson [6] formalized this requirement in four consistency axioms, and then proved that, given new constraint information in terms of expected values, there is only one distribution satisfying these constraints that can be chosen by a procedure that satisfies the consistency axioms; this unique distribution can be obtained by maximizing entropy. After all, inference methods should relate to entropy.

Supposing that the queueing system is in a maximum entropy condition, we obtain the most uncertain probability model compatible with the available information expressed by the moments.

## 2. Maximum Entropy Distribution

The distribution chosen by the maximum entropy principle will be called the maximum entropy (ME) distribution. The density inference problem is formally stated as:

Determine the density $f(x)$ of a random variable $X$ subject to the condition that the expected values $\eta_i$ of $n$ known functions $g_i(X)$ of $X$ are given, that is,

$$E[g_i(X)] = \int_{-\infty}^{\infty} g_i(x) f(x)\, dx = \eta_i, \quad i = 1, \ldots, n. \tag{1}$$

The ME distributions may be easily derived by the following two lemmas. We refer to Papoulis [5] for proofs of them.

**Lemma 1.** The ME distribution must be an exponential form

$$f(x) = A\exp\{-\alpha_1 g_1(x) - \cdots - \alpha_n g_n(x)\}, \tag{2}$$

where $\alpha_i$ are $n$ constants determined from (1) and $A$ is the constant satisfing the density condition

$$A \int_{-\infty}^{\infty} \exp\{-\alpha_1 g_1(x) - \cdots - \alpha_n g_n(x)\} = 1. \tag{3}$$

**Lemma 2.** The following is a system of n equations equivalent to (1) and thus can be used to determine the $n$ parameters $\alpha_i$:

$$-\frac{1}{Z}\frac{\partial Z}{\partial \alpha_i} = \eta_i, \ i = 1, \ldots, n, \tag{4}$$

where the partition function $Z$ is defined as

$$Z = Z(\alpha_1, \ldots, \alpha_n) = \frac{1}{A} = \int_{-\infty}^{\infty} \exp\{-\alpha_1 g_1(x) - \cdots - \alpha_n g_n(x)\} dx. \tag{5}$$

## 3. Some General Results

We will give here some results for a general queueing system.

**Proposition 1.** If the arrival rate or service rate is available, then the ME distribution of the interarrival time or service time is exponential.

**Proof.** We will give a proof for the interarrival time $T$. The same arguments may be applied to the service time $S$. If the arrival rate is $\lambda$, then, by Lemma 1,

$$f_T(t) = Ae^{-\alpha t}, \ 0 < t < \infty. \tag{6}$$

The value of the partition function is

$$Z(\alpha) = \frac{1}{A} = \int_0^{\infty} e^{-\alpha t} dt = \frac{1}{\alpha}, \tag{7}$$

so that

$$-\frac{1}{Z}\frac{\partial Z}{\partial \alpha} = \alpha \frac{1}{\alpha^2} = \frac{1}{\alpha} = \frac{1}{\lambda}. \tag{8}$$

From (7) and (8), we get

$$\alpha = \lambda \text{ and } A = \lambda. \tag{9}$$

Inserting (9) into (6), we have the desired result

$$f_T(t) = \lambda e^{-\lambda t}, \ 0 < t < \infty, \tag{10}$$

which is an exponential distribution with mean $1/\lambda$.

**Proposition 2.** If the expected number of customers in the system is available, then the ME distribution of the system state $N$ is geometric.

**Proof.** If the expected number of customers in the system is $L$, then, by Lemma 1,

$$P_n = Ae^{-\alpha n}, \quad n = 0, 1, \ldots \tag{11}$$

The value of the partition function is

$$Z(\alpha) = \frac{1}{A} = \sum_{n=0}^{\infty} e^{-\alpha n} = \frac{1}{1-e^{-\alpha}}, \tag{12}$$

so that

$$-\frac{1}{Z}\frac{\partial Z}{\partial \alpha} = (1 - e^{-\alpha})\frac{e^{-\alpha}}{(1 - e^{-\alpha})^2} = \frac{e^{-\alpha}}{1 - e^{-\alpha}} = L. \tag{13}$$

From (12) and (13), we get

$$e^{-\alpha} = \frac{L}{1+L} \text{ and } A = \frac{1}{1+L}. \tag{14}$$

Inserting (14) into (11), we have the desired result

$$P_n = P\{N = n\} = \frac{1}{1+L}\left(\frac{L}{1+L}\right)^n, \quad n = 0, 1, \ldots, \tag{15}$$

which is a geometric distribution with the success probability $1/(L+1)$, a discrete analogue of the exponential distribution. See Johnson and Kotz [3] for the distributions.

## 4. Some Specific Results

The Pollaczek-Khintchine formula gives the expected number of customers for an M/G/1 queueing system in a steady-state condition (see Taha [7] p. 619). So we can assert the following proposition for the ME distribution of the system state.

**Theorem 1.** If the service time $S$ in an M/G/1 queueing system in a steady-state condition has the mean $1/\mu$ and the variance $\sigma^2$, then ME distribution of the system state $N$ is given by

$$P_n = P\{N = n\} = \frac{2\mu(\mu-\lambda)\lambda^n(2\mu-\lambda+\lambda\mu^2\sigma^2)^n}{(2\mu^2-\lambda^2+\lambda^2\mu^2\sigma^2)^{n+1}}, \quad n = 0, 1, \ldots, \tag{16}$$

where $\lambda$ is the arrival rate such that $\lambda < \mu$.

**Proof.** The Pollaczek-Khintchine formula gives the expected number of customers $L = E[N]$ for an M/G/1 queueing system in a steady-state condition

$$L = \frac{\lambda}{\mu} + \frac{\lambda^2(1/\mu^2 + \sigma^2)}{2(1-\lambda/\mu)}, \quad \mu > \lambda. \tag{17}$$

Inserting (17) into (15) and applying Proposition 2, we obtain the result.

The classical theory of an M/G/1 queueing system based on a birth-and-death process gives no simple analytical expression for the distribution of the system state, except for the probability of state 0, which is given by

$$\tilde{P}_0 = 1 - \frac{\lambda}{\mu}. \tag{18}$$

On the other hand,

$$P_0 = \frac{1-\lambda/\mu}{1-(1/2)(1-u^2\sigma^2)\lambda^2/\mu^2} = \left[1-\frac{1}{2}\frac{\lambda^2}{\mu^2}(1-\mu^2\sigma^2)\right]^{-1}\tilde{P}_0. \tag{19}$$

If $\sigma^2 = 1/\mu^2$ as is in the case of the exponential service time distributions, then

$$P_0 = \tilde{P}_0 = 1 - \frac{\lambda}{\mu}.$$

From the above theorem for general service time distributions, we can easily deduce the following two corollaries for special service time distributions.

**Corollary 1.** If $\sigma^2 = 1/\mu^2$ as is in the case of the exponential service time distributions, then the ME distribution of the system state $N$ is given by

$$P_n = \left(1-\frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^n, \quad n = 0, 1, \ldots, \tag{20}$$

which is equivalent to the result for an M/M/1 queueing system in a steady-state condition.

**Proof.** If $\sigma^2 = 1/\mu^2$, then the Pollaczek-Khintchine formula (17) reduces to

$$L = \frac{\lambda}{\mu - \lambda}, \tag{21}$$

which is just the ratio between the arrival rate $\lambda$ and the queue decrease rate $\mu - \lambda$. Inserting (21) into (15), we obtain the desired result (20).

**Corollary 2.** If the service time is constant (with $\sigma^2 = 0$), then the ME distribution of the system state $N$ is given by

$$P_n = \frac{2(1-\lambda/\mu)(\lambda/\mu)^n(2-\lambda/\mu)^n}{[2-(\lambda/\mu)^2]^{n+1}}, \quad n = 0, 1, \ldots, \tag{22}$$

which is the result for an M/D/1 queueing system.

**Proof.** If $\sigma^2 = 0$, then the Pollaczek-Khintchine formula (17) reduces to

$$L = \frac{2\lambda/\mu - (\lambda/\mu)^2}{2 - 2\lambda/\mu}. \tag{23}$$

Inserting (23) into (15), we obtain the desired result (22).

**Theorem 2.** If the maximum number of customers allowed in the queueing system is $K$, and if the expected number of customers in the system is $L$, then ME distribution of the system state $N$ is given by

$$P_n = P\{N = n\} = \frac{\delta^n(1-\delta)}{1-\delta^{K+1}}, \quad n = 0, 1, \ldots, K, \tag{24}$$

where $\delta$ is the quantity that has the relationship with $L$

$$L = \frac{\delta[1-(K+1)\delta^K + K\delta^{K+1}]}{(1-\delta^{K+1})(1-\delta)}, \tag{25}$$

which is just the 'classical' result for the M/M/1/K queueing system with $\delta$ playing the role of the traffic intensity $\rho = \lambda/\mu$.

**Proof.** Lemma 1 yields

$$P_n = Ae^{-\alpha n}, \quad n = 0, 1, \ldots, K. \tag{26}$$

The value of the partition function is

$$Z(\alpha) = \frac{1}{A} = \sum_{n=0}^{K} e^{-\alpha n} = \frac{1-(e^{-\alpha})^{K+1}}{1-(e^{-\alpha})} = \frac{1-\delta^{K+1}}{1-\delta}. \tag{27}$$

where

$$\delta = e^{-\alpha}. \tag{28}$$

Thus we have

$$-\frac{1}{Z}\frac{\partial Z}{\partial \alpha} = -\frac{1}{Z}\frac{\partial Z}{\partial \delta}\frac{\partial \delta}{\partial \alpha} = -\frac{1-\delta}{1-\delta^{K+1}}\,\frac{(1-\delta^{K+1})-(K+1)(1-\delta)\delta^n}{(1-\delta)^2}(-\delta) = L,$$

which reduces to the relationship (25). From (27), we get

$$A = \frac{1-\delta}{1-\delta^{K+1}}. \tag{29}$$

Inserting (28) and (29) into (26), we have the desired result (24).

## 5. Further Results

Until now, we have assumed that the available information is represented by the moments of first order. If, however, some moments of higher order are also available, the maximum entropy principle can still be applied.

**Proposition 3.** If both the mean and variance of the interarrival times or service times are available, then the ME distribution of the interarrival time or service time is normal with the given mean and variance.

**Proof.** Here we will give a proof for the interarrival time $T$. The same arguments can also be applied to the service time $S$. If the expected value and variance of the interarrival times are $\frac{1}{\mu_a}$ and $\sigma_a^2$, respectively, then Lemma 1 yields

$$f_T(t) = Ae^{-\alpha_1 t - \alpha_2 t^2}, \quad -\infty < t < \infty. \tag{30}$$

The value of the partition function is

$$Z(\alpha_1, \alpha_2) = \frac{1}{A} = \int_{-\infty}^{\infty} e^{-\alpha_1 t - \alpha_2 t^2} dt = \sqrt{\frac{\pi}{\alpha_2}}\, e^{\frac{\alpha_1^2}{4\alpha_2}}, \tag{31}$$

so that

$$-\frac{1}{Z}\frac{\partial Z}{\partial \alpha_1} = -\frac{\alpha_1}{2\alpha_2} = \frac{1}{\mu_a}, \tag{32}$$

and

$$-\frac{1}{Z}\frac{\partial Z}{\partial \alpha_2} = \frac{1}{2\alpha_2} + \frac{\alpha_1^2}{4\alpha_2^2} = \frac{1}{\mu_a^2} + \sigma_a^2. \tag{33}$$

Equations (32) and (33) yields the solutions

$$\alpha_1 = -\frac{1}{\mu_a \sigma_a^2} \text{ and } \alpha_2 = \frac{1}{2\sigma_a^2}. \tag{34}$$

Inserting (31) and (34) into (30), we have the desired result: $T \sim N\left(\frac{1}{\mu_a}, \sigma_a^2\right)$.

**Proposition 4.** If the mean and variance of the number of customers in the system are $L$ and $\sigma_s^2$, respectively, then the ME distribution of the system state $N$ is

$$P_n = \left(\sum_{k=0}^{\infty} u^k v^{k^2}\right)^{-1} u^n v^{n^2}, \quad n = 0, 1, 2, \ldots, \tag{35}$$

where $u$ and $v$ have the relationships with $L$ and $\sigma_s^2$

$$(\sum_{k=1}^{\infty} nu^n v^{n^2}) / (\sum_{k=0}^{\infty} u^n v^{n^2}) = L, \tag{36}$$

and

$$(\sum_{k=1}^{\infty} n^2 u^n v^{n^2}) / (\sum_{k=0}^{\infty} u^n v^{n^2}) = L^2 + \sigma_s^2. \tag{37}$$

**Proof.** Lemma 1 yields

$$P_n = Ae^{-\alpha_1 n - \alpha_2 n^2}, \quad n = 0, 1, 2, \ldots \tag{38}$$

The value of the partition function is

$$Z(\alpha_1, \alpha_2) = \frac{1}{A} = \sum_{n=0}^{\infty} e^{-\alpha_1 n - \alpha_2 n^2} = 1 + uv + u^2v^4 + u^3v^9 + \cdots, \tag{39}$$

where

$$u = e^{-\alpha_1} \text{ and } v = e^{-\alpha_2}. \tag{40}$$

Thus

$$-\frac{1}{Z}\frac{\partial Z}{\partial \alpha_1} = -\frac{1}{Z}\frac{\partial Z}{\partial u}\frac{\partial u}{\partial \alpha_1} = \frac{uv + 2u^2v^4 + 3u^3v^9 + \cdots}{1 + uv + u^2v^4 + u^3v^9 + \cdots} = L, \tag{41}$$

and

$$-\frac{1}{Z}\frac{\partial Z}{\partial \alpha_2} = -\frac{1}{Z}\frac{\partial Z}{\partial v}\frac{\partial v}{\partial \alpha_2} = \frac{uv + 4u^2v^4 + 9u^3v^9 + \cdots}{1 + uv + u^2v^4 + u^3v^9 + \cdots} = L^2 + \sigma_s^2. \tag{42}$$

Inserting (39) through (42) into (38), we obtain the desired result (35).

## 6. Concluding Remarks

We have attempted to get a probabilistic model of a queueing system in the maximum entropy condition. Applying the maximum entropy principle to the queueing system, we have obtained the most uncertain probability model compatible with the available information expressed by moments. To derive the maximum entropy distribution given second or higher moments may require numerical analysis. We have derived:

- the maximum entropy distributions of interarrival or service time, given the first and/or second moments, and
- the maximum entropy distributions of system states, given the first and/or second moments.

We feel that the present approach may be, at least, a supplementary tool to the classical approach to the queueing theory. If the form of the prior distribution $h$ is available in addition to the moment constraints, the maximum relative entropy principle of Kullback [4] may be applied: Of all the distributions $f$ that satisfy the constraints, you should choose the $f$ with the largest relative entropy $\int f(x) \log (f(x) / h(x)) \, dx$.

## Reference


[1] Guiasu, S. (1986). Maximum Entropy Condition in Queueing Theory, *Journal of Operational Research Society*, Vol. 37, 293-301.

[2] Jaynes, E. T. (1957). Information Theory and Statistical Mechanics I. *Physical Review*. Vol. 106, 620-630.

[3] Johnson, N. L. and Kotz, S. (1969). *Distributions In Statistics: Discrete Distributions*, Houghton Mifflin, Boston.

[4] Kullback, S. (1959). *Information Theory and Statistics*, Wiley, New York.

[5] Papoulis, A. (1991). *Probability, Random Variables, and Stochastic Processes*, 3rd edition, McGraw-Hill, New York.

[6] Shore, J. E. and Johnson, R. W. (1980). Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy, *IEEE Transactions on Information Theory*, Vol. IT-26, 26-37.

[7] Taha, H. A. (1987). *Operations Research*, 4th edition, Macmillan, New York.