The Geometry of M/D/1 Queues and Large Deviation

Kenji NAKAGAWA

Dept. of Electrical Engineering
Nagaoka University of Technology
Nagaoka, Niigata, 940-2188 Japan
Tel. +81-258-21-4263
Fax. +81-258-21-4263
E-mail nakagawa@nagaokaut.ac.jp

Abstract

Since an M/D/1 queue is represented by a Markov chain, we can consider the set of all the M/D/1 queues as a subset of Markov chains. A geometric structures is induced from the geometric structure of the set of Markov chains which forms an exponential family. In this paper, we show that in the large deviation of the tail probability of the queue length of an M/D/1, the rate function and a twisted Markov chain, etc., are represented in terms of the geometry. Moreover, in the importance sampling (IS) simulation for the M/D/1 queue, we elucidate the geometric relation between the underlying distribution and a simulation distribution, and evaluate the variance of an IS estimate by geometric quantities.

Kew words: M/D/1 queue, large deviation, importance sampling, information geometry

1 Introduction

We consider the set of all M/D/1 queues and investigate the geometric properties of the set. Based on the geometric structure, we study the large deviations theory of the tail probability $P(Q > q)$ of the queue length Q in the steady state, and the importance sampling (IS) simulation for $P(Q > q)$.

Amari (1985) introduced the geometric structures into the set of probability distributions, such as a Riemannian metric by the Fisher information and a real one-parameter family of the α-affine connections. Then, statistical quantities are translated into geometric quantities, so a good perspective are given to various statistical problems. There are many applications of the geometric structure, for example, the statistical inference (see Amari, 1985, 1989), hypothesis testing (see Nakagawa, 1993), linear systems (see Amari, 1987), learning theory (see Amari, 1995), and so on, are represented in terms of geometry and many new algorithms have been exploited.

In this paper, we consider the set of M/D/1 queues and introduce a geometric structure on it. An M/D/1 queue is represented as a discrete-time Markov chain, and the set of M/D/1 queues forms a one-dimensional subset of Markov chains with parameter λ which is the arrival rate of a Poisson process. A geometric structure is introduced on the set by the properties of the probability of queue length paths.

In the large deviations theory, for example, as the Sanov theorem states that the rate function is given by the Kullback-Leibler divergence of two probability distributions, we often need to consider two (or more) probability distributions simultaneously (see e.g., Dembo, 1993). Furthermore, as we will show later, in the IS simulation method, many probability distributions appear, such as, the underlying distribution, a simulation distribution, a twisted distribution, etc. Therefore, it is quite natural to consider the relation, especially, the geometric relation between these probability distributions in the set of all probability distributions.

We represent the queue length of M/D/1 by a discrete time Markov chain, and show that the probability of a sample path of the Markov chain forms an exponential family. Next, on the set of M/D/1 queues, we introduce the Riemannian metric, α-affine connections, $+1$-affine coordinate system θ, -1-affine coordinate system η, and the divergence. We show that the rate function and a twisted distribution in the large deviation of M/D/1 queue length are represented by the derived geometry. Moreover, in the IS method, the relation between the underlying distribution and the simulation distribution, and the variance of an IS estimate are evaluated by geometric quantities.

2 Geometric Properties of M/D/1

Let us consider an M/D/1 queue with arrival rate λ and service rate 1. Define a time slot as the service time for one customer. Let Q_n denote the queue length at n th time slot, and A_n the number of arriving customers of the Poisson process at n th time slot. Then the queue length Q_n satisfies the following recursive formula;

$$Q_n = \begin{cases} Q_{n-1} - 1 + A_n, & \text{if } Q_{n-1} > 0, \\ A_n, & \text{if } Q_{n-1} = 0. \end{cases} \tag{1}$$

From (1), we see that $\{Q_n\}$ forms a Markov chain. Write $x_n \equiv Q_n - Q_{n-1}$, $n = 1, 2, \ldots$, then the state transition probability $P(Q_n|Q_{n-1})$ is given by

$$\begin{aligned} P(Q_n|Q_{n-1}) &= P(x_n|Q_{n-1}) \\ &= \begin{cases} \dfrac{\lambda^{x_n+1}}{(x_n+1)!} e^{-\lambda}, & \text{if } Q_{n-1} > 0, \\ \dfrac{\lambda^{x_n}}{x_n!} e^{-\lambda}, & \text{if } Q_{n-1} = 0. \end{cases} \end{aligned} \tag{2}$$

For a path $Q^n \equiv Q_0 Q_1 \ldots Q_n$ of queue lengths, put $x^n \equiv x_1 x_2 \ldots x_n$. We have

$$\begin{aligned} P(Q^n|Q_0) &= P(x^n|Q_0) \\ &= \prod_{i=1}^{n} P(x_i) \\ &= \prod_{i=1}^{n} \frac{\lambda^{x_{i+1}}}{(x_i+1)!} e^{-\lambda}, \end{aligned} \tag{3}$$

for a path with $Q_i > 0,\ i = 0, 1, \ldots$.

Consider the estimate $\bar{x}^n = (1/n) \sum_{i=1}^n x_i$ of the expectation $E[x_i] = \lambda - 1$ of x_i. From (3), we have

$$\frac{1}{n} \log P(x^n|Q_0) = \bar{x}^n \log \lambda - \lambda + \log \lambda - \frac{1}{n} \sum_{i=1}^n \log(x_i + 1)!. \tag{4}$$

Putting $l(x^n) = (1/n) \log P(x^n|Q_0)$, $\theta = \log \lambda$ and $\Psi(\theta) = e^\theta - \theta$, we have

$$l(x^n) = \bar{x}^n \theta - \Psi(\theta) + F(x^n), \tag{5}$$

where $F(x^n)$ is a function of x^n which does not include θ.

Now, denote by $C(\lambda)$ the Markov chain defined by the recursive formula (1) with $0 < \lambda < \infty$ (not restricted as $0 < \lambda < 1$). From (5), we see that the probability of a path generated by $C(\lambda)$ forms an exponential family (see Itoh, 1988) with the natural parameter $\theta = \log \lambda$. Since $C(\lambda)$ is determined by the arrival rate $\lambda > 0$, the set of all $C(\lambda)$ can be identified with the set $\Delta \equiv \mathbb{R}^+ = \{\lambda > 0\}$ of positive real numbers. Based on (5), we can construct a geometric structure of Δ.

First, by regarding $\theta = \log \lambda$ as a coordinate on Δ, a Riemannian metric $g(\theta)$ (see Amari, 1985) on Δ is given by

$$\begin{aligned} g(\theta) &\equiv -E[\frac{\partial^2 l(x^n)}{\partial \theta^2}] \\ &= \lambda. \end{aligned} \tag{6}$$

Next, the α-affine connection $\Gamma^\alpha(\theta)$ (see Amari, 1985) is given by

$$\begin{aligned} &\Gamma^\alpha(\theta) \\ &\equiv E\left[\{\frac{\partial^2 l(x^n)}{\partial \theta^2} + \frac{1-\alpha}{2}(\frac{\partial l(x^n)}{\partial \theta})^2\}\frac{\partial l(x^n)}{\partial \theta}\right] \\ &= \frac{1-\alpha}{2} e^\theta. \end{aligned} \tag{7}$$

For $\alpha = 1$, we have $\Gamma^\alpha(\theta) = \Gamma^{+1}(\theta) = 0$, thus Δ is flat with respect to the $+1$-affine connection, i.e., $+1$ flat. The θ is the $+1$-affine coordinate and $\Psi(\theta) = e^\theta - \theta$ is the potential function (see Amari, 1985).

A $+1$ flat space is simultaneously -1 flat and there exists the -1-affine coordinate η which is dual to θ (Amari, 1985). The η coordinate is given by

$$\begin{aligned} \eta &= \frac{\partial \Psi(\theta)}{\partial \theta} \\ &= \lambda - 1. \end{aligned} \tag{8}$$

Since

$$l(x^n) = \bar{x}^n \log(\eta + 1) - (\eta + 1 - \log(\eta + 1)) + F(x^n)$$

holds, we have

$$
\begin{aligned}
g(\eta) &\equiv -E[\frac{\partial^2 l(x^n)}{\partial \eta^2}] \\
&= \frac{1}{\lambda},
\end{aligned}
\tag{9}
$$

and

$$
\begin{aligned}
\Gamma^{\alpha}(\eta) &\equiv E\left[\{\frac{\partial^2 l(x^n)}{\partial \eta^2} + \frac{1-\alpha}{2}(\frac{\partial l(x^n)}{\partial \eta})^2\}\frac{\partial l(x^n)}{\partial \eta}\right] \\
&= -\frac{1+\alpha}{2}\frac{1}{\lambda^2}.
\end{aligned}
\tag{10}
$$

From $g(\eta) = g(\theta)^{-1}$ and $\Gamma^{-1}(\eta) = 0$, we can confirm that η is the -1-affine coordinate that is dual to θ. The potential function $\Phi(\eta)$ with respect to the η coordinate is given by $\theta = \partial\Phi(\eta)/\partial\eta$, i.e.,

$$
\begin{aligned}
\Phi(\eta) &= \int \log(\eta+1)d\eta \\
&= (\eta+1)\log(\eta+1) - (\eta+1).
\end{aligned}
\tag{11}
$$

Next, the divergence $d(\lambda_1, \lambda_2)$ is defined for $\lambda_1, \lambda_2 \in \Delta$ as follows. Denote by θ_1, θ_2 the θ coordinates of λ_1, λ_2, respectively, and η_1, η_2 the η coordinates of λ_1, λ_2, respectively. Then the divergence $d(\lambda_1, \lambda_2)$ is defined by

$$
\begin{aligned}
d(\lambda_1, \lambda_2) &= \Phi(\eta_1) + \Psi(\theta_2) - \eta_1\theta_2 \\
&= \lambda_1 \log\frac{\lambda_1}{\lambda_2} - \lambda_1 + \lambda_2.
\end{aligned}
$$

The following lemma holds.

Lemma 1 *For $\lambda_1,\ \lambda_2 \in \Delta$, $d(\lambda_1, \lambda_2) \geq 0$ holds and the equality holds if and only if $\lambda_1 = \lambda_2$.*

In summary, we have

Theorem 1 *The set Δ of the Markov chains that are defined by (1) has the structure of ± 1 flat manifold with $+1$-affine coordinate $\theta = \log\lambda$, and -1-affine coordinate $\eta = \lambda - 1$. The divergence $d(\lambda_1, \lambda_2)$ for $\lambda_1, \lambda_2 \in \Delta$ is given by*

$$
d(\lambda_1, \lambda_2) = \lambda_1 \log\frac{\lambda_1}{\lambda_2} - \lambda_1 + \lambda_2. \tag{12}
$$

Among the θ and η coordinates introduced above, we especially take the θ coordinate into consideration. A function θ_t of a real parameter t which satisfies the differential equation

$$
\frac{d^2\theta_t}{dt^2} + \Gamma^{+1}(\theta_t)(\frac{d\theta_t}{dt})^2 = 0
$$

is called a $+1$ geodesic. Since $\Gamma^{+1}(\theta) = 0$, the $+1$ geodesic is a straight line with respect to the θ coordinate, i.e., θ_t can be represented as

$$\theta_t = t\theta_1 + \theta_2 \tag{13}$$

for some θ_1, θ_2. Here, we must notice that since Δ is a one-dimensional set, so any curves which pass λ_1, λ_2 are identical as a set of points, however, only in the case that θ_t is parametrized by (13) we call it a $+1$ geodesic.

The $+1$ geodesic has a close relation with a twisted distribution which will be mentioned below.

3 Large Deviation of M/D/1

Let us denote by Q the queue length in the steady state of M/D/1 queue with an arrival rate λ. Consider the large deviation for the tail probability of Q.

Rewrite the recursive formula (1) by replacing Q_n by f_n to have

$$f_n = f_{n-1} + V(f_{n-1}, A_n), \tag{14}$$

where A_n is the number of arrivals of the Poisson process with arrival rate λ and

$$V(f_{n-1}, A_n) = \begin{cases} A_n - 1, & \text{if } f_{n-1} > 0, \\ A_n, & \text{if } f_{n-1} = 0. \end{cases} \tag{15}$$

Denote by $C(\lambda)$ the Markov chain defined by (14) with $0 < \lambda < \infty$. If $f_{n-1} > 0$, then for the jump size $V \equiv V(f_{n-1}, A_n)$, the expectation $E[V]$, the moment generating function $M(t)$, and the rate function $I(x)$ are given by

$$\begin{aligned} E[V] &= \lambda - 1 && (16) \\ M(t) &= \exp(\lambda e^t - t - \lambda) && (17) \\ I(x) &\equiv \sup_{t\in\mathbb{R}}(tx - \log M(t)) && (18) \\ &= (x+1)\log\frac{x+1}{\lambda} - (x+1) + \lambda. && (19) \end{aligned}$$

By scaling the time axis, jump size, and the arrival rate, we approximate f_n by a differentiable path $f(\tau)$ (Bucklew, 1990). By (14) and the law of large numbers, the average path $\bar{f}_\lambda(\tau)$ of a Markov chain $C(\lambda)$ satisfies

$$\bar{f}'_\lambda(\tau) = E[V] = \lambda - 1.$$

Therefore, we have

$$\bar{f}_\lambda(\tau) = (\lambda - 1)\tau. \tag{20}$$

Namely, the probability model of paths is determined by the arrival rate $\lambda > 0$, and the average path (20) is the straight line with slope $\lambda - 1$. The slope $\lambda - 1$ coincides with the η coordinate $\eta = \lambda - 1$ on Δ.

In the large deviations theory, the rate function is determined by a distance-like measure of two probability models. For an arbitrary path $f(\tau)$, $0 \leq \tau \leq T$, of a Markov chain $C(\lambda)$, the rate function $R(f)$ of $f(\tau)$ is given (Bucklew, 1990) by

$$R(f) = \int_0^T I(f'(\tau))d\tau.$$

Especially, for the average path $\bar{f}_{\widetilde{\lambda}}(\tau)$ of $C(\widetilde{\lambda})$ with an arrival rate $\widetilde{\lambda}$, we have

$$\begin{aligned} R(\bar{f}_{\widetilde{\lambda}}) &= T(\widetilde{\lambda} \log \frac{\widetilde{\lambda}}{\lambda} - \widetilde{\lambda} + \lambda) \\ &= Td(\widetilde{\lambda}, \lambda). \end{aligned} \tag{21}$$

From (21), we see that the rate function of an average path can represented by the divergence.

Our goal is to have an approximating value of $P(Q > q)$ of the queue length Q in the steady state. According to the large deviations theory, the exponential decay rate of $P(Q > q)$ is determined by the rate function of the average path $\bar{f}_{\widetilde{\lambda}}$ with $\widetilde{\lambda}$ that minimizes the $R(\bar{f}_{\widetilde{\lambda}})$. T is the time that the path $\bar{f}_{\widetilde{\lambda}}$ reaches at 1, i.e., we have $T = 1/(\widetilde{\lambda} - 1)$ (Bucklew, 1990). Thus, we have

$$R(\bar{f}_{\widetilde{\lambda}}) = \frac{d(\widetilde{\lambda}, \lambda)}{\widetilde{\lambda} - 1}. \tag{22}$$

Since the average path $\bar{f}_{\widetilde{\lambda}}$ exceeds 1 if and only if $\widetilde{\lambda} > 1$, (22) should be minimized with respect to $\widetilde{\lambda} > 1$. In fact, by differentiating (22) with $\widetilde{\lambda}$ and putting the derivative equal to 0, we have

$$-\log \frac{\widetilde{\lambda}}{\lambda} - \lambda + \widetilde{\lambda} = 0. \tag{23}$$

For the solution $\widetilde{\lambda} = \lambda^*$ of (23), the $C(\lambda^*)$ is the dominating Markov chain, and the probability $P(Q > q)$ is approximated (Bucklew, 1990) as

$$P(Q > q) \approx e^{-R(\bar{f}_{\lambda^*})q}. \tag{24}$$

Now, we investigate the rate function $R(\bar{f}_{\lambda^*})$ in (24) from the view point of geometry.

First of all, we define a twisted distribution for $C(\lambda)$. Recall that the probability of a path of $C(\lambda)$ is determined by the probability distribution of V which is the difference of queue lengths between two consecutive time slots, i.e.,

$$P_\lambda(V = b) \equiv \frac{\lambda^{b+1}}{(b+1)!} e^{-\lambda}, \; b = -1, 0, 1, \ldots. \tag{25}$$

Then, for the distribution (25), a *twisted distribution* $P_{\lambda,t}$ is defined by

$$P_{\lambda,t}(V=b) \equiv \frac{e^{tb}}{M(t)} \frac{\lambda^{b+1}}{(b+1)!} e^{-\lambda}, \ \ b=-1,0,1,\dots,$$

where $M(t)$ is the moment generating function defined by (17).

By a simple calculation, we have

$$P_{\lambda,t} = \frac{(\lambda e^t)^{b+1}}{(b+1)!} e^{-\lambda e^t}. \tag{26}$$

Hence, the twisted distribution $P_{\lambda,t}$ for P_λ is equal to $P_{\lambda e^t}$.

Write $\lambda_t \equiv \lambda e^t$ and denote by $\theta_\lambda = \log \lambda$ and $\theta_{\lambda,t} = \log \lambda_t$ the θ coordinates of λ and λ_t, respectively, then we have

$$\theta_{\lambda,t} = t + \theta_\lambda. \tag{27}$$

So, from (13), we see that the twisted distribution $P_{\lambda,t}$ is a $+1$ geodesic. By substituting $\widetilde{\lambda} = \lambda_t = \lambda e^t$ into (23), we have a equation

$$\lambda e^t - t - \lambda = 0, \tag{28}$$

and for the solution $t = t^*$ to (28), we have from (12) and (22),

$$R(\bar{f}_{\lambda_{t^*}}) = t^*. \tag{29}$$

We can summarize the above as

Theorem 2 *The large deviation rate function of the queue length of M/D/1 is a value of the parameter t of a $+1$ geodesic on Δ whose value is determined by* (28).

4 Importance Sampling Simulation of M/D/1

We apply the importance sampling (IS) simulation method [5] to the estimation of the tail probability $P(Q > q)$. IS method is one of the fast simulation techniques, in which a larger arrival rate γ than the underlying rate λ is used to generate customers more frequently. As a result, the queue length becomes longer so the target event $\{Q > q\}$ can be observed more frequently. Thus the simulation is accelerated.

If the arrival rate γ used in the simulation is chosen appropriately, then the variance of the IS estimate can be smaller than that of the ordinary Monte Carlo (MC) method. In particular, the probability distribution that minimizes the variance of the IS estimate is called the *optimal* simulation distribution, and the arrival rate of the optimal simulation distribution is called the optimal arrival rate.

The principal problem in the IS method is to obtain the optimal simulation distribution. In the case of the tail probability of the queue length in the steady state of

M/D/1, the optimal arrival rate γ is determined as follows. The average path, which exceeds q, of the Markov chain $C(\gamma)$ with the optimal γ should be given the largest probability with respect to the underlying distribution among the average paths of all the Markov chains. As discussed in the previous section, the path with the largest probability, or the dominating path is the average path $\bar{f}_{\lambda^*}$ with the arrival rate λ^* which is the solution to (23). Hence the Markov chain P_{λ^*} is the optimal simulation distribution.

Thus, we have

Theorem 3 [6] *The optimal arrival rate λ^* of the IS simulation for the tail probability of the M/D/1 queue length is given by $\lambda^* = \lambda e^{t^*}$, where t^* is the solution to the equation*

$$\lambda e^t - t - \lambda = 0.$$

4.1 Variance of IS Estimate

Let us denote by $\sigma^2_{\gamma,q}$ the variance of an IS estimate for $P(Q > q)$ where γ is the arrival rate of a simulation distribution. The asymptotic rate t_γ of the variance is defined by

$$t_\gamma \equiv \lim_{q\to\infty} -\frac{1}{q} \log \sigma^2_{\gamma,q}. \tag{30}$$

The following result is known.

Theorem 4 [12] *For the underlying arrival rate λ and the arrival rate γ of a simulation distribution, the asymptotic rate t_γ satisfies the equation*

$$\frac{\lambda^2}{\gamma} e^t - t - 2\lambda + \gamma = 0.$$

The inequality $t_\gamma \le 2t^*$ holds and the equality holds if and only if $\gamma = \lambda^*$, where λ^* is given by Theorem 3.

Now, we can represent those fundamental properties described above in terms of the information geometry (see Nakagawa, 1996, 2001). For preparation, as an analogue of the twisted distribution defined in the previous section, we here define a γ-twisted distribution for the arrival rate γ of a simulation distribution.

A *γ-twisted distribution*, which is denoted by $P_{\lambda^2/\gamma,t}(b)$, is defined by

$$P_{\lambda^2/\gamma,t}(b) = \frac{e^{tb}}{M_\gamma(t)} \frac{(\lambda^2/\gamma)^{b+1}}{(b+1)!} e^{-\lambda^2/\gamma}, \; b = -1, 0, 1, \dots, \; t \in \mathbb{R}, \tag{31}$$

where $M_\gamma(t)$ is the moment generating function of the probability distribution

$$P_{\lambda^2/\gamma}(b) = \frac{(\lambda^2/\gamma)^{b+1}}{(b+1)!} e^{-\lambda^2/\gamma}.$$

By a simple calculation, we have

$$M_\gamma(t) \equiv \exp(\lambda^2 e^t/\gamma - t - \lambda^2/\gamma),$$

and, as a consequence,

$$P_{\lambda^2/\gamma,t}(b) = \frac{(\lambda^2 e^t/\gamma)^{b+1}}{(b+1)!} e^{-\lambda^2/\gamma}. \tag{32}$$

We see that a γ-twisted distribution is $P_{\lambda^2 e^t/\gamma}$.

We have considered so far the underlying distribution P_λ, a twisted distribution $P_{\lambda,t}$, and a simulation distribution P_γ with arrival rate γ. Furthermore, we defined in (31) a γ-twisted distribution $P_{\lambda^2/\gamma,t}$. The geometric relation between these four distributions is presented and the variance of the IS estimate is represented by the geometric quantities.

Let us denote by $\theta_\lambda,\ \theta_{\lambda,t},\ \theta_\gamma$, and $\theta_{\lambda^2/\gamma,t}$ the θ coordinates of probability distributions $P_\lambda,\ P_{\lambda,t},\ P_\gamma$, and $P_{\lambda^2/\gamma,t}$, respectively. By definition, we have

$$\theta_\lambda = \log\lambda \tag{33}$$

$$\theta_{\lambda,t} = t + \log\lambda \tag{34}$$

$$\theta_\gamma = \log\gamma \tag{35}$$

$$\theta_{\lambda^2/\gamma,t} = t + 2\log\lambda - \log\gamma \tag{36}$$

We have

Theorem 5

$$\theta_{\lambda^2/\gamma,t} - \theta_\lambda = \theta_{\lambda,t} - \theta_\gamma. \tag{37}$$

Proof: It is easily shown by (33)-(36).

The relation (37) of the θ coordinates can be depicted as in Fig.1.

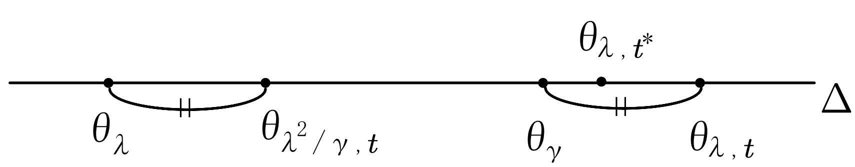

Figure 1: Relation of $P_\lambda,\ P_{\lambda,t},\ P_\gamma$, and $P_{\lambda^2/\gamma,t}$ in θ coordinate.

For the underlying arrival rate λ, let us consider the optimal IS arrival rate $\lambda^* = \lambda e^{t^*}$ which is determined by Theorem 3, and the arrival rate $\lambda^2 e^{t_\gamma}/\gamma$ of the γ-twisted distribution with t_γ which is determined by Theorem 4. For these arrival rates, we have

Theorem 6

$$2t^* - t_\gamma = \frac{d(\lambda^*, \gamma) + d(\lambda^*, \lambda^2 e^{t_\gamma}/\gamma)}{\lambda^* - 1}.$$

Proof: By a straightforward calculation.

From Theorem 6, Lemma 1, and $\lambda^* > 1$, we see that $t_\gamma \leq 2t^*$ holds and the equality holds if and only if the arrival rate γ of the IS simulation is equal to the optimal rate λ^*.

5 Conclusion

We studied the large deviations theory for the tail probability $P(Q > q)$ of the queue length Q in the steady state from the view point of information geometry. We introduced a geometric structure into the set of all M/D/1 queues. Based on the geometric structure, we showed that the rate function of $P(Q > q)$, the dominating distribution, the optimal IS simulation distribution and the variance of the IS estimates, and so on, can be represented by geometric quantities, such as θ coordinate, geodesic, or divergence. Due to the geometric representation, the proofs of the above theorems becomes almost trivial.

Further study is necessary for the geometric structure of more complex queueing models.

References

[1] Amari, S., 1985. Differential-Geometrical Methods in Statistics (Lecture Notes in Statistics), New York: Springer-Verlarg.

[2] Amari, S., 1987. Differential Geometry of a Parametric Family of Invertible Linear Systems — Riemannian Metric, Dual Affine Connections, and Divergence, Math. Systems Theory 20, pp.53-82.

[3] Amari, S., Han, T.S., 1989. Statistical Inference under Multiterminal Rate restrictions: A differential Geometric Approach, IEEE Trans. Inform. Theory, vol.35, no.2, pp.217-227, March.

[4] Amari, S., 1995. Information Geometry of the EM and em Algorithms for Neural Networks, Neural Networks, vol.8, no.9, pp.1379-1408.

[5] Bucklew, J.A., 1990. Large Deviation Techniques in Decision, Simulation, and Estimation, Wiley, New York.

[6] Cottrell, M., Fort, J-C., Malgouyres, G., 1983. Large Deviations and Rare Events in the Study of Stochastic Algorithms, IEEE Trans. Automat. Contr. vol.AC-28, no.9, pp.907-920, September.

[7] Dembo, A., Zeitouni, O., 1993. Large Deviations Techniques and Application, Jones and Bartlett.

[8] Itoh, H., Amari, S., 1988. Geometry of Information Sources, Proc. of the 11th SITA, pp.57-60.

[9] Nakagawa, K., Kanaya, F., 1993. On the Converse Theorem in Statistical Hypothesis Testing for Markov Chains, IEEE Trans. Inform. Theory, vol.39, no.2, pp.629-633, March.

[10] Nakagawa, K., 1996. On the Twisted Markov Chain of Importance Sampling Simulation, IEICE Trans. Fund., vol.E79-A, no.9, pp.1423-1428, September.

[11] Nakagawa, K., 2001. On the Optimal Markov Chain of IS Simulation, IEEE Trans. Inform. Theory, vol.47, no.1, pp.442-446, January.

[12] Sadowsky, J.S., 1991. Large Deviations Theory and Efficient Simulation of Excessive Backlogs in $GI/GI/m$ Queues, IEEE Trans. Automat. Contr. vol.36, no.12, pp.1383-1394, December.