Bayesian Forecasting of WWW Traffic on the Time Varying Poisson Model

Daiki Koizumi[1], **Toshiyasu Matsushima**[2], **and Shigeichi Hirasawa**[1]
[1]Research Institute for Science and Engineering, Waseda University, Tokyo, Japan
[2]Department of Applied Mathematics, Waseda University, Tokyo, Japan

Abstract— *Traffic forecasting from past observed traffic data with small calculation complexity is one of important problems for planning of servers and networks. Focusing on World Wide Web (WWW) traffic as fundamental investigation, this paper would deal with Bayesian forecasting of network traffic on the time varying Poisson model from a viewpoint from statistical decision theory. Under this model, we would show that the estimated forecasting value is obtained by simple arithmetic calculation and expresses real WWW traffic well from both theoretical and empirical points of view.*

Keywords: World Wide Web (WWW) traffic, traffic engineering, statistical decision theory, time varying Poisson distribution, long-range dependence (LRD)

1. Introduction

Under network environment such as Internet, planning of servers and networks is one of important problems for stable operation. It is often typical situation that administrators analyze logs on their servers and networks. They may frequently look into result of log analysis software where these tools usually have some functions to periodically summarize logs. For example, webalizer [1] and analog [9] etc. have been widely used among World Wide Web (WWW) server administrators or users for long years. These tools usually summarize the logs by counting hourly, daily, and monthly numbers of hits, files, and pages etc. Administrators would often make their operation plans with combination of their experience and intuition from these logs. In this case, traffic forecasting rule is not clearly formulated and those summarized logs remain in the field of *descriptive statistics* from the statistical point of view.

On the other hand, researchers in the field of traffic engineering have been suggesting a lot of analysis models. Probabilistic approach is one of viewpoints in this field. It is wide-spread fact that the stationary Poisson distribution is not always suitable for Internet traffic because of its nature of non-stationality [6] [4] and long-range dependence (LRD) [5] [6] etc. Therefore desirable conditions of good traffic models are to have structures to express such nature at least. Furthermore, another requirement of models is to have a structure of traffic forecasting. For this point, parameter estimation is often performed at first under assumption of the stationarity [7], then the estimated parameter is substituted for the parameter of model. This approach has been wide-spread in the field of *inferential statistics* from the statistical point of view.

However, substituting the estimated parameter as a constant for the model's parameter is not always suitable especially on forecasting problems. This is because there is often no guarantee that the assumptions under the parameter estimation of the model always hold for future unknown data set. Bayesian approach [2][3] is one of alternatives for this point. In Bayesian approach, a probability distribution of parameter is assumed as the prior distribution. If new data is observed, then the Bayes theorem updates the prior distribution of parameter to the posterior distribution and then forecasts the posterior distribution of data. Recently, this approach has been widely applied to many forecasting problems especially in the field of information technologies and bioinformatics etc. In order to take Bayesian approach, *statistical decision theory* is an important theoretical framework from the statistical point of view.

Taking the above factors into account, this paper would deal with Bayesian forecasting of WWW traffic on the non-stationary i.e. time varying Poisson model. Bayesian forecasting on time varying parameter model has been proposed in [8] by defining certain class of parameter transformation function. However, it has not yet been discussed about any predictive estimator nor definite transformation function of parameter [8]. This paper would clearly define a random-walking type of transformation function of parameter to obtain the Bayes optimal prediction for WWW traffic. Then its effectiveness would be evaluated with real WWW traffic data. In this model, time varying degree is caught by a real valued constant $k\,(0 < k \leq 1)$ and this constant would play an important role throughout this paper. Another feature is that the traffic forecasting value is obtained by simple arithmetic calculations under known k. In general, the Bayes theorem often results in large calculation costs. However, certain combination of parameter distribution and its transformation function solves this problem. We believe that this point can be helpful not only for theoretical calculation cost but also for real implementation on WWW log analysis tools [1] [9].

The rest of this paper is organized as the followings. Section 2 gives some definitions and explanations of the forecasting model with time varying Poisson distribution. Section 3 shows some analysis examples of real WWW traffic data to validate this paper's approach and Section 4 gives their discussions. Finally, Section 5 concludes this paper.

2. The Time Varying Poisson Model

2.1 Definitions

Suppose X is a discrete random variable and $p(x) = Pr\{X = x\}$ is the probability distribution where real number x is each element of space of X. Throughout this paper, the probability distribution of X is assumed to depend on real valued parameter $\theta \in \Theta$. The true parameter $\theta^* \in \Theta$ is unknown, however, the probability distribution of parameter $p(\theta)$ is assumed to be known. Hereafter the probability distribution of X under parameter θ is simply denoted as $p(x \mid \theta)$.

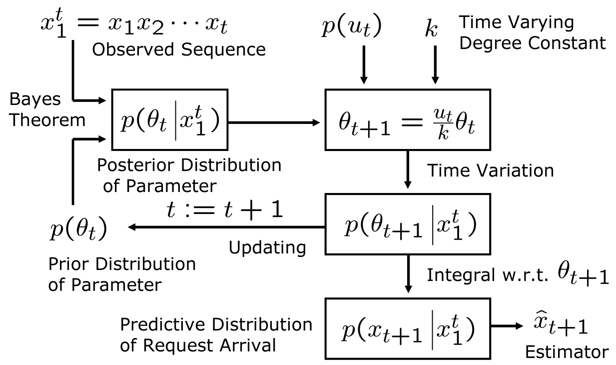

Fig. 1: Overview of the inferential process.

Let $t = 1, 2, \cdots$ be discrete time and $x_t = 0, 1, \cdots$ be number of WWW request arrivals at time t, respectively. This paper focuses on x_t for traffic analysis by assuming probability distribution $p(x_t \mid \theta_t)$ where $\theta_t > 0$ is a time varying density parameter at time t. The time varying Poisson model takes sequence of $x_1^t = x_1 x_2 \cdots x_t$ as input and calculates $\hat{x}_{t+1}$ as an output estimator where the prior distribution of parameter $p(\theta_t)$ and time variation rule of θ_t are known. The overview of the inferential process is depicted in Figure 1.

In Figure 1, x_t is assumed to be the Poisson distribution with a time varying density parameter θ_t as follows:

For $x_t = 0, 1, 2, \cdots$,

$$p(x_t \mid \theta_t) = \frac{\exp(-\theta_t)}{x_t!} (\theta_t)^{x_t}, \tag{1}$$

where $\theta_t > 0$ is a time varying density parameter.

For parameter θ_t, the following time varying model is assumed:

For $\theta_{t+1}, \theta_t > 0$,

$$\theta_{t+1} = \frac{u_t}{k}\theta_t, \tag{2}$$

where k is a constant such that $0 < k \leq 1$, and $0 < u_t < 1$ is a continuous random variable which is conditionally independent from θ_t. (2) represents a transformation of θ_{t+1} from both θ_t and u_t under a known constant k. This transformation is regarded as a kind of random-walk.

Furthermore, the initial random variables of θ_1 and u_1 are from the Gamma and Beta distributions, respectively. The followings give their definitions:

For $\theta_1 > 0$,

$$p(\theta_1 \mid \alpha_1, \beta_1) = \frac{(\beta_1)^{\alpha_1}}{\Gamma(\alpha_1)} \exp(-\theta_1 \beta_1) (\theta_1)^{(\alpha_1)-1}, \tag{3}$$

where $\alpha_1 > 0,\ \beta_1 > 0$ are parameters of the Gamma distribution.

In (3), $\Gamma(x)$ is the Gamma function defined below:

$$\Gamma(x) = \int_0^\infty y^{x-1} \exp(-y)\, dy, \tag{4}$$

where $x > 0$.

For $0 < u_1 < 1$,

$$p(u_1 | k\alpha_1, (1-k)\alpha_1) = \frac{\Gamma(\alpha_1)}{\Gamma(k\alpha_1)\Gamma[(1-k)\alpha_1]} (u_1)^{(k\alpha_1)-1} (1-u_1)^{[(1-k)\alpha_1]-1}, \tag{5}$$

where $k\alpha_1 > 0,\ (1-k)\alpha_1 > 0$ are parameters of the Beta distribution.

In (3) and (5), two random variables θ_1 and u_1 have a real value $\alpha_1 > 0$ as their parameters in common. This means that θ_1 and u_1 are conditionally independent.

Remarks 2.1: In (2), a constant $0 < k \leq 1$ expresses time varying degree of θ_t. If $k = 1$, (2) simply becomes $\theta_{t+1} = u_t \theta_t$. In this case, θ_{t+1} does not vary since the variance of u_t , which equals to $k(1-k)/(\alpha_t+1)$ according to the nature of Beta distribution, becomes zero. This means that the Poisson distribution of x_t in (1) is stationary.

If $k < 1$, on the other hand, θ_{t+1} varies depending on the previous θ_t which expresses the time varying Poisson model. If $k = 0.5$, the time varying degree of θ_t becomes maximum since the variance of u_t takes the maximum value. Thus the proposed model defined in (1)–(5) includes a classical stationary Poisson distribution as a special case if $k = 1$. Moreover, k plays another important role which would be explained in *Remarks 2.2*.

2.2 Updating Density Parameter θ_t

As shown in Figure 1, the parameter updating rule from θ_t to θ_{t+1} consists of Bayes theorem and time variation. In the followings, this subsection 2.2 is divided into three parts. The first 2.2.1 describes the updating rule by Bayes theorem for $t \geq 2$, the second 2.2.2 describes the initial condition of the prior distribution $p(\theta_1)$ for $t=1$, and the last 2.2.3 describes general time variation rule which is mainly discussed in [8].

2.2.1 The Posterior Distribution of Density Parameter

Suppose that sequence x_1^{t-1} is already observed where $t \geq 2$ and the prior distribution of parameter $p\left(\theta_t \mid \alpha_t, \beta_t, x_1^{t-1}\right)$ is defined as the Gamma distribution in (3) for $t \geq 2$. If new x_t is observed, the posterior distribution of $p\left(\theta_t \mid x_1^t\right)$ by the Bayes theorem is obtained as the following:

$$p\left(\theta_t \mid x_1^t\right) = \frac{p(x_t \mid \theta_t) p(\theta_t \mid \alpha_t, \beta_t, x_1^{t-1})}{\int_0^\infty p(x_t \mid \theta_t) p(\theta_t \mid \alpha_t, \beta_t, x_1^{t-1}) d\theta_t} \quad (6)$$

$$= \frac{(\beta_t+1)^{\alpha_t+x_t}}{\Gamma(\alpha_t+x_t)} \exp[-\theta_t(\beta_t+1)](\theta_t)^{(\alpha_t+x_t)-1}, \quad (7)$$

where $t \geq 2$.

(7) means that the posterior distribution $p\left(\theta_t \mid x_1^t\right)$ is also Gamma with parameter $(\alpha_t + x_t)$ and $(\beta_t + 1)$.

2.2.2 Initial Condition of Prior Distribution of Density Parameter

For the initial prior distribution of $p(\theta_1)$, this paper assumes no anomalies for WWW traffic. In Bayesian context, *non-informative prior* [2][3] can be considered as the following:

$$p(\theta_1) \propto \frac{1}{\theta_1}. \quad (8)$$

The above distribution can be formulated by the Gamma distribution with parameters $\alpha_1 > 0, \beta_1 = 1$ in (3). Therefore, the general posterior updating form in (7) can also be applied for $t = 1$. Thus (7) holds for any $t \geq 1$. This means that the posterior distribution $p\left(\theta_t \mid x_1^t\right)$ can be simply calculated for any $t \geq 1$ by considering two parameters α_t, β_t on the Gamma distribution if the initial prior distribution of (8) is assumed. This is the nature of *conjugate prior* [2] [3] between the Poisson and Gamma distributions. This nature contributes drastic reduction of calculation complexity under large t.

2.2.3 Time Variation of Density Parameter

To obtain $p\left(\theta_{t+1} \mid x_1^t\right)$, a time variation of density parameter defined in (2) is used. This is actually a transformation of random variables among u_t, θ_t, and constant k. Even if a transformation is newly defined after Bayes theorem, the distribution family of density parameter remains same as that of the conjugate prior distribution under certain class of transformations. Such class has been discussed in [8] as the following.

Theorem 1 (Simple Power Steady Model [8]): Suppose a parameter distribution $p(\theta_t)$ is in the linear expanding family. Let T be a transformation function of parameter θ_t. If T satisfies the following condition, then the parameter distribution remains same family of distribution not depending on t and the forecasting model is called *Simple Power Steady Model (S.P.S.M.)* with respect to T:

$$T(x) = \Psi x^k,\ \Psi > 0,\ 0 < k < 1. \quad (9)$$

If T satisfies (9) and is one-to-one mapping, the model can also be S.P.S.M.[8]. However, it has not yet been discussed about any definite transformation function which is required to obtain the forecasting estimator. In fact, the time varying parameter model defined in (3) is one-to-one mapping since Jacobian of (3) is easily proved to be non-zero. Thus the model in this paper is actually included in S.P.S.M. The forecasting estimator would be derived in the next subsection 2.3.

Under S.P.S.M. in this paper, the transformed distribution of θ_{t+1} now becomes:

$$p\left(\theta_{t+1} \mid x_1^t\right) = p\left(\frac{u_t}{k}\theta_t \mid x_1^t\right) \quad (10)$$

$$= \frac{[k(\beta_t+1)]^{k(\alpha_t+x_t)}}{\Gamma[k(\alpha_t+x_t)]} \times \exp[-\theta_{t+1}k(\beta_t+1)](\theta_{t+1})^{k(\alpha_t+x_t)-1}. \quad (11)$$

(10) and (11) mean that the transformed distribution of θ_{t+1} becomes the Gamma distribution with the following parameters:

$$\begin{cases} \alpha_{t+1} = k(\alpha_t + x_t); \\ \beta_{t+1} = k(\beta_t + 1). \end{cases} \quad (12)$$

If (12) is recursively applied with respect to t, the following equations are obtained:

$$\begin{cases} \alpha_{t+1} = k^t\alpha_1 + \sum_{i=1}^t k^{t+1-i} x_i; \\ \beta_{t+1} = k^t\beta_1 + \sum_{i=1}^t k^{i-1}. \end{cases} \quad (13)$$

The above equations contribute drastic reduction of calculation complexity.

2.3 Output Estimator $\hat{x}_{t+1}$

The output of proposed model is an estimator $\hat{x}_{t+1}$ as depicted in Figure 1. $\hat{x}_{t+1}$ is a prediction of number of request arrivals at time $(t+1)$ under the input sequence $x_1^t = x_1 x_2 \cdots x_t$. $\hat{x}_{t+1}$ is formulated in terms of statistical decision theory [2][3] and derived as the following.

Let $\hat{x}_{t+1}$ be an estimator of x_{t+1} and define the following squared-error loss function to evaluate $\hat{x}_{t+1}$:

$$L(\hat{x}_{t+1}, x_{t+1}) = (\hat{x}_{t+1} - x_{t+1})^2. \quad (14)$$

Since x_{t+1} distributes with the Poisson defined in (1), *risk function* [2][3] under the certain density parameter θ_{t+1} becomes the following:

$$R(\hat{x}_{t+1}, \theta_{t+1}) = \sum_{x_{t+1}=0}^{\infty} \left[L(\hat{x}_{t+1}, x_{t+1})\, p(x_{t+1} \mid \theta_{t+1})\right] \quad (15)$$

$$= \sum_{x_{t+1}=0}^{\infty} \left[(\hat{x}_{t+1} - x_{t+1})^2\, p(x_{t+1} \mid \theta_{t+1})\right]. \quad (16)$$

Next, *the Bayes risk function* [2][3], which is obtained by taking expectation of the risk function with respect to density parameter θ_{t+1}, becomes the following:

$$BR(\hat{x}_{t+1})$$
$$=\int_0^\infty R(\hat{x}_{t+1},\theta_{t+1})p(\theta_{t+1})d\theta_{t+1} \tag{17}$$
$$=\sum_{x_{t+1}=0}^{\infty}(\hat{x}_{t+1}-x_{t+1})^2\int_0^\infty p(x_{t+1}|\,\theta_{t+1})p(\theta_{t+1})d\theta_{t+1}\,. \tag{18}$$

Finally, suppose an estimator $\hat{x}^*_{t+1}$ to minimize the Bayes risk function defined in (18). Such estimator is called *the Bayes optimal prediction* [2][3]. Under this paper's assumptions, $\hat{x}^*_{t+1}$ is obtained as follows:

$$\hat{x}^*_{t+1}$$
$$=\arg\min_{\hat{x}_{t+1}} BR\left(\hat{x}_{t+1}\right) \tag{19}$$
$$=\sum_{x_{t+1}=0}^{\infty} x_{t+1}\int_0^\infty p(x_{t+1}\mid\theta_{t+1},x_1^t)p(\theta_{t+1}|x_1^t)d\theta_{t+1} \tag{20}$$
$$=E\left[x_{t+1}\mid x_1^t\right] \tag{21}$$
$$=E\left[\theta_{t+1}\mid x_1^t\right] \tag{22}$$
$$=\frac{\alpha_{t+1}}{\beta_{t+1}} \tag{23}$$
$$=\frac{k^t\alpha_1+\sum_{i=1}^t k^{t+1-i}x_i}{k^t\beta_1+\sum_{i=1}^t k^{i-1}}. \tag{24}$$

Note that (22) is obtained since x_{t+1} is the Poisson distribution defined in (1) and the expectation of x_{t+1} corresponds to the that of density parameter θ_{t+1}. (23) is obtained since θ_{t+1} has the Gamma distribution defined in (3) and its expectation becomes α_{t+1}/β_{t+1}. (24) is obtained by applying (13) to (23).

Remarks 2.2: In (24), $\hat{x}^*_{t+1}$ is obtained by simple arithmetic calculation. This point can be effective not only theoretical point of view but also the real implementation such as server log analysis software tools. The second term of numerator in (24) has a form of *Exponentially Weighted Moving Average* [8] with a time varying constant k. As k becomes larger in (24), the weighting of past observed sequence x_1^t increases. This means that k can be considered as a parameter of long-range dependence (LRD). If $k=1$, its weighting becomes maximum and the model is regarded as stationary model. The effect of k under the real WWW traffic data would be considered in the next section.

3. Analysis Examples of WWW Traffic Data

3.1 WWW Traffic Data

The real WWW traffic data was derived from access logs of two different WWW servers (A, and B) on campus. The number of request arrivals was counted with every five-minutes-interval except for maintenance period. The detail specifications are described in Table 1.

Table 1: Specifications of real WWW traffic data.

	Server A	Server B
Request Arrivals	154,932	274,302
Start Date	Mar. 18, 2005	Mar. 18, 2006
End Date	Apr. 9, 2005	Apr. 9, 2006
Time Length	13d 12h 10m	13d 1h 10m
Time Intervals	3,890	3,758

3.2 Maximum Likelihood Estimation for k

Properties described in the previous section hold on condition that a time varying constant k in (2) is known. If real data is dealt with, however, k is unknown and should be estimated. Taking the maximum likelihood estimation of k, the objective likelihood function $L(k)$ becomes the following:

$$L(k)$$
$$=p(x_1\mid\theta_1)\prod_{i=2}^t p(x_i\mid x_1^{i-1},k) \tag{25}$$
$$=p(x_1\mid\theta_1)\prod_{i=2}^t\left[\int_0^\infty p(x_i|\,x_1^{i-1},\theta_i)p(\theta_i\,|\,x_1^{i-1})d\theta_i\right] \tag{26}$$
$$=p(x_1\mid\theta_1)$$
$$\times\prod_{i=2}^t\left[\frac{(\beta_i)^{\alpha_i}\Gamma(\alpha_i+x_i)}{(\beta_i+1)^{\alpha_i+x_i}\Gamma(\alpha_i)x_i!}\,\middle|\,\begin{matrix}\alpha_i=k^{i-1}\alpha_1+\sum_{j=1}^{i-1}k^{i-j}x_j\\ \beta_i=k^{i-1}\beta_1+\sum_{j=1}^{i-1}k^{j-1}\end{matrix}\right]. \tag{27}$$

Note that α_i,β_i in (27), the previously obtained results in (13) were applied. Therefore, the maximum likelihood estimator (MLE) of $\hat{k}$ is obtained as follows:

$$\hat{k}=\arg\max_k L\left(k\right). \tag{28}$$

On the above likelihood function, the analytical maximum likelihood estimation for k is quite difficult, however, the solution can be obtained by numerical calculation. The interval $0\le k\le 1$ is divided into 1,000 sub-intervals and value of $\log L(k)$ is calculated for each k numerically. Some plots of function $\log L(k)$ are shown in Figure 2. Some examples of MLE for k are also shown in Table 2.

3.3 Point Estimation for WWW Traffic Forecasting

The real WWW data described on Table 1 was processed to evaluate the point estimates of future request arrivals on the proposed model. For the performance comparison, the point estimates on the classical stationary Poisson model were also calculated. On the proposed model, each MLE

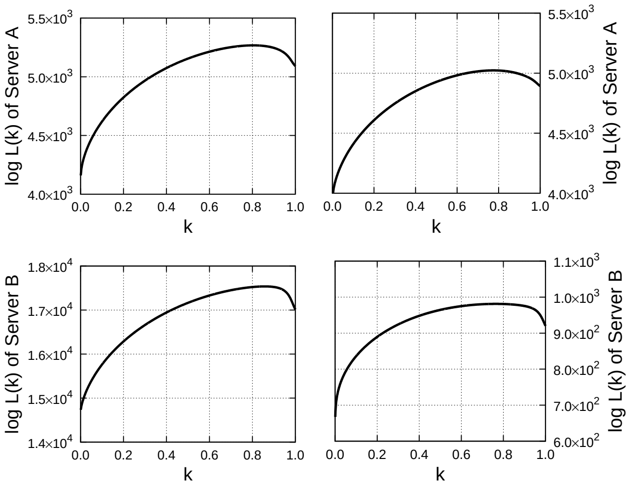

Fig. 2: Examples of log-likelihood functions. Left top is in Mar. 23, 2005 on server A; Right top is in Mar. 31, 2005 on server A; Left bottom is in Mar. 27, 2006 on server B; Right bottom is in Apr. 08, 2006 on server B.

Table 2: Maximum likelihood estimates for k.

Name	Date	MLE of $\hat{k}$
Server A	Mar. 23, 2005	0.804
Server A	Mar. 31, 2005	0.775
Server B	Mar. 27, 2006	0.857
Server B	Apr. 08, 2006	0.764

of $\hat{k}$ was calculated from the previous day's log. To evaluate performance on both models, the mean squared error between each point estimate and observed value of request arrivals was calculated.

Table 3 and 4 show mean squared error of proposed and stationary models on server A and B, respectively. In each table, the MLEs of $\hat{k}$ from the previous day's logs are showed on the third row. Figure 3 shows point and interval estimates v.s. observed values plot of server A on Mar. 25, 2005 where $\hat{k} = 0.804$. In Figure 3, the vertical axis is the number of request arrivals and the horizontal axis is time interval index. The solid line, solid chain line, and histogram represent the point estimates on the proposed model, those on the classical stationary Poisson model, and observed values of request arrivals, respectively. Dotted lines are 95% interval estimates of proposed and stationary models. Figure 4 also shows point and interval estimates v.s. observed values plot of server B on Mar. 28, 2006 where $\hat{k} = 0.857$.

3.4 Interval Estimation for WWW Traffic Forecasting

Table 5 shows interval estimation example of server A on Mar. 25, 2005. In Table 5, $t = 104$ is taken since it

Table 3: Mean squared error on server A.

Server A	Proposed Model	$\hat{k}$	Stationary Model
Mar. 19	2.829×10^2	0.716	4.988×10^2
Mar. 20	2.657×10^2	0.762	3.297×10^2
Mar. 21	3.816×10^2	0.753	4.802×10^2
Mar. 22	7.111×10^2	0.805	9.534×10^2
Mar. 23	8.202×10^2	0.759	1.335×10^3
Mar. 24	1.356×10^3	0.804	3.458×10^3
Mar. 25	9.523×10^2	0.804	2.062×10^3
Mar. 26	4.811×10^2	0.783	6.479×10^2
Mar. 27	8.596×10^2	0.771	1.239×10^3
Mar. 28	1.980×10^3	0.754	4.041×10^3
Mar. 29	1.657×10^3	0.777	4.019×10^3
Mar. 30	4.940×10^2	0.788	7.568×10^2
Mar. 31	8.088×10^2	0.787	1.218×10^3
Apr. 01	8.967×10^2	0.775	1.887×10^3
Apr. 02	1.258×10^1	0.753	1.220×10^1
Apr. 03	4.184×10^0	0.826	4.375×10^0
Apr. 04	4.206×10^1	0.914	4.317×10^1
Apr. 05	4.095×10^1	0.666	3.808×10^1
Apr. 06	3.612×10^1	0.710	4.723×10^1
Apr. 07	2.813×10^2	0.661	5.183×10^2
Apr. 08	2.295×10^1	0.786	2.325×10^1
Apr. 09	7.589×10^0	0.803	8.127×10^0

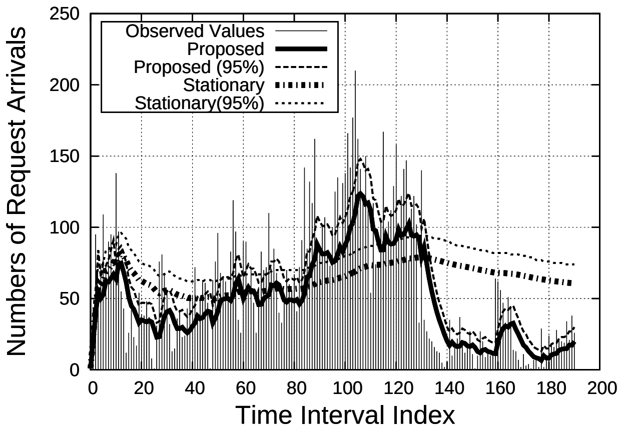

Fig. 3: Point and interval estimates v.s. observed values plot of server A on Mar. 25, 2005 ($\hat{k} = 0.804$).

gives max observed value $x_{104} = 210$ as the numbers of request arrivals. The second, third, and forth rows show the expected value, 95% confidence limit, and 99% confidence limit, respectively on the proposed and stationary models. Table 6 also shows interval estimation result of server B on Mar. 28, 2006 where $t = 86$ is taken.

4. Discussion

4.1 Maximum Likelihood Estimation for k

According to Figure 2, it is showed that there exist some cases where their likelihood functions of $\log L(k)$ are convex. Actually, all likelihood functions were convex to the

Table 4: Mean squared error on server B.

Server B	Proposed Model	$\hat{k}$	Stationary Model
Mar. 19	3.070×10^2	0.777	3.874×10^2
Mar. 20	1.302×10^3	0.820	2.386×10^3
Mar. 21	5.159×10^2	0.782	5.896×10^2
Mar. 22	1.244×10^3	0.832	1.650×10^3
Mar. 23	1.741×10^3	0.827	4.151×10^3
Mar. 24	1.109×10^3	0.836	1.980×10^3
Mar. 25	4.705×10^2	0.812	5.436×10^2
Mar. 26	1.504×10^3	0.755	1.915×10^3
Mar. 27	2.964×10^3	0.800	7.918×10^3
Mar. 28	1.109×10^1	0.857	1.932×10^1
Mar. 29	2.019×10^3	0.805	5.576×10^3
Mar. 30	4.370×10^2	0.784	6.708×10^2
Mar. 31	5.318×10^2	0.799	6.750×10^2
Apr. 01	2.568×10^2	0.710	3.327×10^2
Apr. 02	4.739×10^2	0.640	5.118×10^2
Apr. 03	6.019×10^2	0.745	8.237×10^2
Apr. 04	1.544×10^3	0.791	5.448×10^3
Apr. 05	1.449×10^2	0.820	1.709×10^2
Apr. 06	4.712×10^2	0.791	5.620×10^2
Apr. 07	2.784×10^2	0.777	3.567×10^2
Apr. 08	3.200×10^3	0.764	1.082×10^4
Apr. 09	1.333×10^3	0.841	3.128×10^3

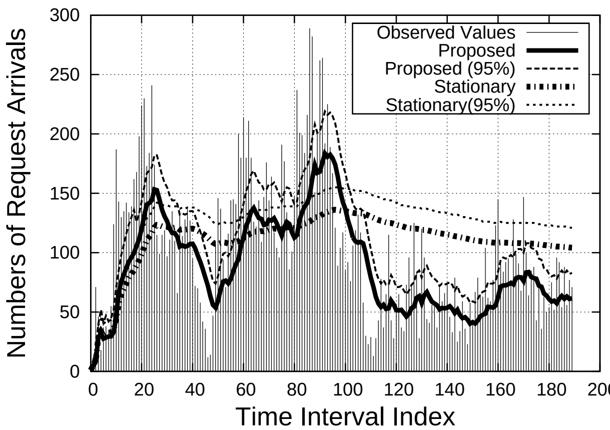

Fig. 4: Point and interval estimates v.s. observed values plot of server B on Mar. 28, 2006 ($\hat{k} = 0.857$).

best of numerical calculations in this paper. Figure 2 also shows that the absolute value of gradient in $\log L(k)$ around MLE of $\hat{k}$ becomes quickly larger as k increases beyond $\hat{k}$. This fact shows that the under estimation for k causes less error than its over estimation.

For the estimated value of k, the minimum was $\hat{k} = 0.640$ and the maximum was $\hat{k} = 0.914$ as shown in Table 2, 3, and 4. Most of $\hat{k}$ distributes between 0.700 and 0.850. This fact suggests non-stationarity of the WWW traffic data. Table 7 shows Akaike Information Criterion (AIC) on server A. According to Table 7, most of AIC on the proposed model are smaller than those of stationary model to select the proposed model. Exception is that absolute values of AIC

Table 5: Interval estimation of server A on Mar. 25, 2005.

$t = 104$ on Server A	Proposed	Stationary
$\hat{x}_{104}$ (Expected Value)	111	69
$\hat{x}_{104}$ (95% Confidence Limit)	133	83
$\hat{x}_{104}$ (99% Confidence Limit)	142	89
x_{104} (Max Observed Value)	210	

Table 6: Interval estimation of server B on Mar. 28, 2006.

$t = 86$ on Server B	Proposed	Stationary
$\hat{x}_{86}$ (Expected Value)	148	126
$\hat{x}_{86}$ (95% Confidence Limit)	170	145
$\hat{x}_{86}$ (99% Confidence Limit)	180	153
x_{86} (Max Observed Value)	289	

on the two models suddenly become smaller around Apr. 01, 2005 on server A. This is actually because the average traffic on server A drastically decreased to less than 1,000 request arrivals per a day. In this period, the traffic in each time interval stayed at lower level and could be regarded as the stationary Poisson model. For server B, on the other hand, such traffic decrease did not occur and the proposed model is chosen for all days by AIC model selection. As a whole, it can be concluded that the proposed model has stronger validity for real WWW traffic data than the stationary model from the viewpoint of model selection.

4.2 Point Estimation for WWW Traffic Forecasting

For point estimation of future request arrivals, Table 3, and 4 show that the proposed model has the better performance than that of the stationary model in terms of mean squared error. Figure 3 and 4 also depict that the point estimates on the proposed model are following more closely to the observed values than those on stationary model. As mentioned in *Remarks 2.1*, the proposed model contains the stationary model as a special case when $k = 1.000$. Therefore, regardless of its stationarity or non-stationarity of WWW traffic, the proposed model can be applied to traffic forecasting and would help for planning of WWW servers to some extent.

However, it should be noted that this result strongly depends on the accuracy of MLE of k. In Table 3, each MLE of k during days in April often differs from $\hat{k} = 1.000$ in spite of its strong stationarity on the real traffic. In such situation, the mean squared error on the proposed model becomes larger than that of stationary model. Another example is that if $k = 0.300$ on the proposed model with server A, the mean squared error of the proposed model becomes 5.41×10^3 where that of stationary model becomes 3.46×10^3. Figure 5 depicts this poor performance of the proposed model. Figure 6 is a plot of mean squared errors v.s. k. In interval of $k < 0.600$, the extremely smaller estimate of k could cause larger mean squared error. The under estimation near the MSE minimizer of k, however, causes relatively smaller error than its over estimation as previously described in subsection 4.1.

In Figure 6, mean squared errors takes minimum values around $k = 0.800$. In fact, Table 3 and 4 show that corresponding maximum likelihood estimates for k are $k = 0.783$ for server A and $k = 0.805$ for server B, respectively. This result suggests that the data length of previous day's log at the maximum likelihood estimation for k was sufficient.

4.3 Interval Estimation for WWW Traffic Forecasting

For interval estimation, time interval indices that give the maximum number of request arrivals are taken on Table 5 and 6. This is because one of administrators' concerns can be the maximum number of the request arrivals in terms of stable server operations. As a result, each confidence limit of x_t derives larger value than that of point estimate of x_t (=expected value) and reduces the mean squared error than that of point estimate. This effect on the proposed model would be stronger than that on the stationary model, since the performance of point estimates of x_t on the proposed model is superior to that of stationary model. Thus the advantage of Bayesian approach was observed.

Table 7: Akaike Information Criterion (AIC) on server A.

AIC on Server A	Proposed	Stationary
Mar. 19, 2005	-3.562×10^3	-3.367×10^3
Mar. 20, 2005	-3.102×10^3	-3.011×10^3
Mar. 21, 2005	-5.085×10^3	-4.960×10^3
Mar. 22, 2005	-8.316×10^3	-8.108×10^3
Mar. 23, 2005	-1.052×10^4	-1.018×10^4
Mar. 24, 2005	-1.797×10^4	-1.715×10^4
Mar. 25, 2005	-1.088×10^4	-1.031×10^4
Mar. 26, 2005	-3.917×10^3	-4.649×10^3
Mar. 27, 2005	-9.023×10^3	-8.723×10^3
Mar. 28, 2005	-1.875×10^4	-1.795×10^4
Mar. 29, 2005	-1.869×10^4	-1.774×10^4
Mar. 30, 2005	-5.825×10^3	-5.610×10^3
Mar. 31, 2005	-1.003×10^4	-9.784×10^3
Apr. 01, 2005	-6.456×10^3	-5.783×10^3
Apr. 02, 2005	-1.538×10^1	-2.313×10^1
Apr. 03, 2005	$+1.763 \times 10^1$	$+8.860 \times 10^0$
Apr. 04, 2005	-1.655×10^2	-1.615×10^2
Apr. 05, 2005	-1.821×10^2	-1.803×10^2
Apr. 06, 2005	-1.771×10^2	-1.516×10^2
Apr. 07, 2005	-2.826×10^3	-2.675×10^3
Apr. 08, 2005	-1.170×10^2	-1.184×10^2
Apr. 09, 2005	$+6.273 \times 10^0$	$+4.767 \times 10^{-1}$

Table 8: Akaike Information Criterion (AIC) on server B.

AIC on Server B	Proposed	Stationary
Mar. 19, 2006	-3.270×10^3	-3.195×10^3
Mar. 20, 2006	-1.244×10^4	-1.201×10^4
Mar. 21, 2006	-1.245×10^4	-1.201×10^4
Mar. 22, 2006	-1.652×10^4	-1.634×10^4
Mar. 23, 2006	-2.231×10^4	-2.165×10^4
Mar. 24, 2006	-1.217×10^4	-1.180×10^4
Mar. 25, 2006	-4.482×10^3	-4.395×10^3
Mar. 26, 2006	$+1.504 \times 10^3$	$+1.915 \times 10^3$
Mar. 27, 2006	-3.504×10^4	-3.403×10^4
Mar. 28, 2006	-2.130×10^4	-2.086×10^4
Mar. 29, 2006	-1.958×10^4	-1.852×10^4
Mar. 30, 2006	-4.601×10^3	-4.428×10^3
Mar. 31, 2006	-3.417×10^3	-3.282×10^3
Apr. 01, 2006	-8.741×10^2	-8.145×10^2
Apr. 02, 2006	-4.393×10^3	-4.308×10^3
Apr. 03, 2006	-7.633×10^3	-7.462×10^3
Apr. 04, 2006	-1.837×10^4	-1.710×10^4
Apr. 05, 2006	-1.033×10^3	-9.828×10^2
Apr. 06, 2006	-3.311×10^3	-3.183×10^3
Apr. 07, 2006	-1.958×10^3	-1.838×10^3
Apr. 08, 2006	-3.384×10^4	-3.211×10^4
Apr. 09, 2006	-1.090×10^4	-1.010×10^4

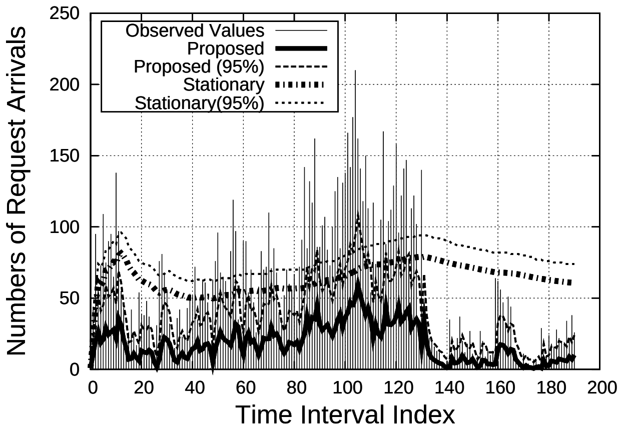

Fig. 5: Point and interval estimates v.s. observed values plot of server A on Mar. 25, 2005 ($k = 0.300$).

5. Conclusion

This paper showed Bayesian forecasting of WWW traffic on the time varying Poisson model. This model is obtained by defining a random-walk type of time varying parameter function on Simple Power Steady Model. The forecasting estimator of this model guarantees the Bayes optimality in terms of statistical decision theory and is calculated by simple arithmetic calculation. The latter point especially can be effective for the real implementation such as server log analysis software tools.

Furthermore, the non-stationarity is expressed by a time varying degree constant k in the model. This paper pointed out that the constant k can be considered as a parameter of long-range dependent (LRD) for real traffic data and the model includes stationary Poisson model as a special case if $k = 1$.

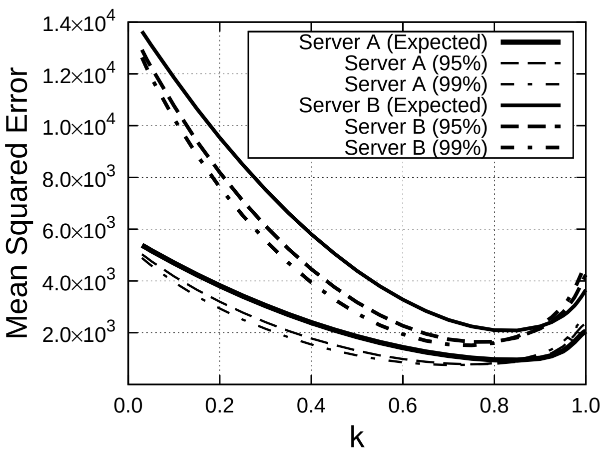

Fig. 6: Mean squared error v.s. k plot of server A on Mar. 25, 2005 and server B on Mar. 28, 2006.

For evaluation of Bayesian approach, the real WWW traffic data is applied to the model in this paper. The maximum likelihood estimation method of k from real traffic data is also discussed and its performance is proved to be sufficient for WWW traffic forecasting.

According to its result, the proposed model has stronger validity than classical stationary Poisson model in terms of model selection. Furthermore, under the estimated value of k, the point and interval estimates on the proposed model showed smaller mean squared error comparing to those on the stationary model for the traffic forecasting. Thus the advantage of the proposed model is shown from both theoretical and empirical points of view.

Acknowledgment

The first and second authors are supported by the following research grants:

- Grant-in-Aid for Scientific Research on Innovative Areas, Japan Society for the Promotion of Science (No.20200044)
- Waseda University Grant for Special Research Projects (No.2009A-058)

References

[1] Bradford L. Barrett, The Webalizer. [Online]. http://www.webalizer.org/

[2] James O.Berger, Statistical Decision Theory and Bayesian Analysis, Springer-Verlag, New York, 1985.

[3] Jose M. Bernardo and Adrian F. M. Smith, Bayesian Theory, John Wiley & Sons, Chichester, 2003.

[4] T. Karagiannis, M. Molle, M. Faloutsos, A. Broido, “A Nonstational Poisson View of Internet Traffic, ” Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2004), Mar. 2004.

[5] W.E.Leland, M.S.Taqqu, “On the Self-Similar Nature of Ethernet Traffic,” SIGCOMM’93, Sep. 1993.

[6] Vern Paxon and Sally Floyd, “Wide Area Traffic: The Failure of Poisson Modeling, ” IEEE/ACM Trans. on Networking, vol.3, no.3, Jun. 1995.

[7] Antoine Scherrer, Nicolas Larrieu, Philippe Owezarski, Pierre Borgnat, Patrice Abry, “Non-Gaussian and Long Memory Statistical Characterizations for Internet Traffic with Anomalies,” IEEE Trans. on Dependable and Secure Computing, vol.4, no.1, pp.56–70, 2007.

[8] J. Q. Smith, “A Generalization of the Bayesian Steady Forecasting Model, ” Journal of the Royal Statistical Society, Series B, vol.41, pp.375–387, 1979.

[9] Stephen Turner, Analog. [Online]. http://www.analog.cx/