Credible Interval Prediction of a Nonstationary Poisson Distribution based on Bayes Decision Theory

Daiki Koizumi[a]

Otaru University of Commerce, 3–5–21, Midori, Otaru-city, Hokkaido, 045–8501, Japan

Keywords: Bayes Decision Theory, Credible Interval, Nonstationary Poisson Distribution.

Abstract: A credible interval prediction problem of a nonstationary Poisson distribution in terms of Bayes decision theory is considered. This is the two-dimensional optimization problem of the Bayes risk function with respect to two variables: upper and lower limits of credible interval prediction. We prove that these limits can be uniquely obtained as the upper or lower percentile points of the predictive distribution under a certain loss function. By applying this approach, the Bayes optimal prediction algorithm for the credible interval is proposed. Using real web traffic data, the performance of the proposed algorithm is evaluated by comparison with the stationary Poisson distribution.

1 INTRODUCTION

The credible interval (or credibility interval) (Berger, 1985; Press, 2003) is a Bayesian interval estimation method defined by a set function with a simpler definition compared to the confidence interval. Since it is a more general estimation method than the Bayesian point estimation, a specific approach (Winkler, 1972) in terms of Bayes decision theory (Weiss and Blackwell, 1961; Berger, 1985) is proposed. In this approach, the Bayesian credible interval parameter estimation method is considered on the assumption of a certain loss function. Specifically, assuming power loss functions, the necessary and sufficient conditions to obtain a unique credible interval solution as the minimizer of the posterior expected loss is discussed.

On the other hand, the Poisson distribution is a well-known probability mass function in various fields. Especially if counting data with a probability model is considered, the Poisson distribution is one important choice. In basic modeling, the stationary Poisson distribution is often defined, which means that its parameter is time independent. However, for a certain types of counting data such as web traffic, the stationary Poisson distribution can be insufficient. Accordingly, the author previously proposed a new class of nonstationary Poisson distribution (Koizumi et al., 2009). In this nonstationary Poisson distribution, its parameter was time dependent and changing with random walking. This nonstationary class was defined as a transforming function of a random variable with a single hyper parameter. Then, assuming the squared loss function to measure the predictive error, the Bayes optimal predictive point estimator in terms of Bayes decision theory was obtained. This estimator can be calculated with simple arithmetic operations under both a certain assumption for the prior distribution of the parameter and the known value of the nonstationary single hyper parameter. Furthermore, this estimator enables the online prediction algorithm and its predictive error (mean squared error, MSE) from real web traffic data was shown to be smaller than that of the stationary Poisson distribution.

In fact, the above Bayes optimal point predictive estimator was defined as the expectation of the predictive distribution. If the squared loss function is defined, then this is natural in terms of Bayes decision theory. In a statistical sense, the expectation can be interpreted as the "central value" of the probability distribution. However, in some fields like web traffic analysis, system operators may be concerned with not the central value but the "upper value" of the request arrival. In order to discuss those upper values of the distribution, the credible interval with a set function can be a useful definition in the context of Bayesian statistics. Furthermore, the credible interval estimation is the generalization of point estimation. These points are expected to enable the credible interval prediction algorithm to be proposed using the

[a] https://orcid.org/0000-0002-5302-5346

Koizumi, D.
Credible Interval Prediction of a Nonstationary Poisson Distribution based on Bayes Decision Theory.
DOI: 10.5220/0009182209951002
In *Proceedings of the 12th International Conference on Agents and Artificial Intelligence (ICAART 2020) - Volume 2*, pages 995-1002
ISBN: 978-989-758-395-7; ISSN: 2184-433X

author's nonstationary Poisson distribution. This paper discusses these points.

The remainder of this paper is organized as follows. Section 2 provides the basic definitions of the nonstationary Poisson distribution and two theorems in terms of Bayesian statistics. Section 2 also describes the conventional results for the credible interval parameter estimation problem in terms of Bayes decision theory. Section 3 formulates the credible interval prediction problem, derives a theorem in terms of Bayes decision theory, and proposes the prediction algorithm. Section 4 gives some numerical examples with real web traffic data. Section 5 discusses the results. Section 6 draws some conclusions of this paper.

2 PRELIMINARIES

2.1 Web Traffic Modeling with Nonstationary Poisson Distribution

Let $t = 1, 2, \dots$ be a discrete time index and $X_t = x_t \geq 0$ be a discrete random variable at t. Assume that web Traffic at time is X_t and $X_t \sim Poisson(\theta_t)$, where $\theta_t > 0$, is a nonstationary parameter. Thus, the probability density function of the nonstationary Poisson distribution $p(x_t \mid \theta_t)$ is defined as follows:

Definition 2.1. *Nonstationary Poisson Distribution*

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

where $\theta_t > 0$. □

A nonstationary class of parameters θ_t is defined as random walking:

Definition 2.2. *Nonstationary Class of Parameter*

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

where $0 < k \leq 1, 0 < u_t < 1$. □

In Eq. (2), a real number $0 < k \leq 1$ is a known constant, $U_t = u_t$ is a continuous random variable, where $0 < u_t < 1$. The probability distribution of u_t is defined in Definition 2.5.

The parameter $\Theta_t = \theta_t$ is a continuous random variable from a Bayesian viewpoint. The prior $\Theta_1 \sim Gamma(\alpha_1, \beta_1)$, where $\theta_1 > 0, \alpha_1 > 0$, and $\beta_1 > 0$. This prior distribution is defined as follows:

Definition 2.3. *Prior Gamma Distribution for* θ_1

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

where $\alpha_1 > 0, \beta_1 > 0$ *and* $\Gamma(\cdot)$ *is the gamma function defined in Definition 2.4.* □

Definition 2.4. *Gamma Function*

$$\Gamma(a) = \int_0^\infty b^{a-1} \exp(-b) db, \tag{4}$$

where $b \geq 0$. □

$\forall t,\ U_t \sim Beta[k\alpha_t, (1-k)\alpha_t]$, where $0 < u_t < 1, 0 < k \leq 1$, and $\alpha_t > 0$. Its probability density function is defined as follows:

Definition 2.5. *Beta Distribution for* u_t

$$p[u_t \mid k\alpha_t, (1-k)\alpha_t] = \frac{\Gamma(\alpha_t)}{\Gamma(k\alpha_t)\Gamma[(1-k)\alpha_t]} (u_t)^{k\alpha_t - 1} (1-u_t)^{(1-k)\alpha_t - 1}. \tag{5}$$

□

Random variables θ_t, u_t are conditional independent under α_t. This is defined as follows:

Definition 2.6. *Conditional Independence for* θ_t, u_t *under* α_t

$$p(\theta_t, u_t \mid \alpha_t) = p(\theta_t \mid \alpha_t) p(u_t \mid \alpha_t). \tag{6}$$

□

Let $\boldsymbol{x}^{t-1} = (x_1, x_2, \dots, x_{t-1})$ be the observed data sequence. Then, the posterior distribution $p(\theta_t \mid \alpha_t, \beta_t, \boldsymbol{x}^{t-1})$ can be obtained with the following closed form.

Theorem 2.1. *Posterior Distribution of* θ_t

$\forall t \geq 2,\ \Theta_t \mid \boldsymbol{x}^{t-1} \sim Gamma(\alpha_t, \beta_t)$. *This means that the posterior distribution* $p(\theta_t \mid \alpha_t, \beta_t, \boldsymbol{x}^{t-1})$ *satisfies the following:*

$$p\left(\theta_t \mid \alpha_t, \beta_t, \boldsymbol{x}^{t-1}\right) = \frac{(\beta_t)^{\alpha_t}}{\Gamma(\alpha_t)} (\theta_t)^{\alpha_t - 1} \exp(-\beta_t \theta_t), \tag{7}$$

where its parameters α_t, β_t *are given as,*

$$\begin{cases} \alpha_t = k^{t-1}\alpha_1 + \sum_{i=1}^{t-1} k^{t-i} x_i; \\ \beta_t = k^{t-1}\beta_1 + \sum_{i=1}^{t-1} k^{i-1}. \end{cases} \tag{8}$$

□

Proof of Theorem 2.1.
See APPENDIX A. □

Theorem 2.2. *Predictive Distribution of* x_{t+1}

$$p(x_{t+1} \mid \boldsymbol{x}^t) = \frac{\Gamma(\alpha_{t+1} + x_{t+1})}{x_{t+1}!\,\Gamma(\alpha_{t+1})} \left(\frac{\beta_{t+1}}{\beta_{t+1}+1}\right)^{\alpha_{t+1}} \left(\frac{1}{\beta_{t+1}+1}\right)^{x_{t+1}}, \tag{9}$$

where $\alpha_{t+1}, \beta_{t+1}$ *are given as Eqs. (9).* □

Proof of Theorem 2.2.
See APPENDIX B. □

2.2 Credible Interval Parameter Estimation based on Bayes Decision Theory

Interval estimation is defined by a set function in contrast to point estimation defined by single value. This interval is particularly called as *credible interval* in terms of Bayesian method. Let $C = [a,b], a \leq b$ be a credible interval. Then, $100(1-\lambda)\%$ credible interval for θ_t can be defined as follows:

Definition 2.7. *Credible Interval Parameter Estimation for* θ_t *(Berger, 1985)*

$$1-\lambda \leq p\left(C \mid \boldsymbol{x}^t\right) = \int_a^b p\left(\theta_t \mid \boldsymbol{x}^t\right) d\theta_t\,, \tag{10}$$

where $0 \leq \lambda < 1$. □

Definition 2.8. *Loss Function for Credible Interval Estimation (Winkler, 1972)*

$$L_1(a,b,\theta_t) = \begin{cases} L_u(\theta_t - b) + r(b-a)\,, & \text{if } b \leq \theta_t\,; \\ r(b-a)\,, & \text{if } a \leq \theta_t \leq b\,; \\ L_o(a-\theta_t) + r(b-a)\,, & \text{if } \theta_t \leq a\,, \end{cases} \tag{11}$$

where $r > 0$ *and,* L_u, L_o *are monotone nondecreasing function with* $L_o(x) = L_u(x) = 0$ *for all* $x \leq 0$. □

Definition 2.9. *Expected Loss (Winkler, 1972)*

$$\begin{aligned} EL(a,b) &= \int_0^{+\infty} L_1(a,b,\theta_t)\, p\left(\theta_t \mid \boldsymbol{x}^t\right) d\theta_t \\ &= \int_0^a L_o(a-\theta_t)\, p\left(\theta_t \mid \boldsymbol{x}^t\right) d\theta_t \\ &\quad + \int_b^{+\infty} L_u(\theta_t - b)\, p\left(\theta_t \mid \boldsymbol{x}^t\right) d\theta_t + r(a-b)\,. \end{aligned} \tag{12}$$

□

With the above definitions, the Bayes optimal credible interval $[\hat{a}, \hat{b}]$ for θ_t is obtained as follows:

$$\hat{a} = \arg\min_a EL(a,b)\,, \tag{13}$$

$$\hat{b} = \arg\min_b EL(a,b)\,. \tag{14}$$

Thus the credible interval with decision theoretic approach is obtained as the two dimensional optimization problem. However, this problem does not always guarantee the unique optimal solution. To do so, more restrictive conditions are needed. The following two propositions solve this problem.

Proposition 2.1. *(Winkler, 1972)*

If L_o, L_u *are convex and not everywhere constant,* $EL(a,b)$ *is finite for all* (a,b)*, then* $EL(a,b)$ *has a minimum value and the set of optimal intervals* (a,b) *is a bounded convex set. If, in addition, either* L_o *or* L_u *is strictly convex, the optimal interval is unique.* □

Proposition 2.2. *(Winkler, 1972)*

If L_o, L_u *are twice differentiable on* $[0,\infty)$ *and the optimal interval has non zero length, then necessary first- and second-order conditions for* (a,b) *to be optimal are,*

$$\int_0^a L_o'(a-\theta_t)\, p\left(\theta_t \mid \boldsymbol{x}^t\right) d\theta_t = \int_b^{+\infty} L_u'(\theta - b)\, p\left(\theta_t \mid \boldsymbol{x}^t\right) d\theta_t = r\,, \tag{15}$$

$$\int_0^a L_o''(a-\theta_t)\, p\left(\theta_t \mid \boldsymbol{x}^t\right) d\theta_t + L_o'(0)\, p\left(a \mid \boldsymbol{x}^t\right) \geq 0\,, \tag{16}$$

$$\int_b^{+\infty} L_u''(\theta_t - b)\, p\left(\theta_t \mid \boldsymbol{x}^t\right) d\theta_t + L_u'(0)\, p\left(b \mid \boldsymbol{x}^t\right) \geq 0\,. \tag{17}$$

□

Lemma 2.1. *(Winkler, 1972)*

Let denote the loss function as the following power function,

$$L_2(a,b,\theta_t) = \begin{cases} c_1(\theta_t - b)^q + c_3(b-a)\,, & \text{if } b \leq \theta_t\,; \\ c_3(b-a)\,, & \text{if } a \leq \theta_t \leq b\,; \\ c_2(a-\theta_t)^r + c_3(b-a)\,, & \text{if } \theta_t \leq a\,, \end{cases} \tag{18}$$

where $c_1, c_2, c_3, q, r > 0$,

then, the necessary first-order condition corresponding Eq. (15) is,

$$\begin{aligned} qc_1 \int_b^{+\infty} (\theta_t - b)^{q-1}\, p\left(\theta_t \mid \boldsymbol{x}^t\right) d\theta &= rc_2 \int_0^a (a-\theta_t)^{r-1}\, p\left(\theta_t \mid \boldsymbol{x}^t\right) d\theta \\ &= c_3\,. \end{aligned} \tag{19}$$

□

Lemma 2.2. *(Winkler, 1972)*

If $q = r = 1$ *in Eq. (18), then the first-order necessary condition is,*

$$\int_0^a p\left(\theta_t \mid \boldsymbol{x}^t\right) d\theta_t = \frac{c_3}{c_2}\,, \tag{20}$$

$$\int_b^{+\infty} p\left(\theta_t \mid \boldsymbol{x}^t\right) d\theta_t = \frac{c_3}{c_1}\,. \tag{21}$$

Furthermore, if $(c_3/c_1) + (c_3/c_2) < 1$*, then the Bayes optimal solution* $\left[\hat{a}, \hat{b}\right]$, $\hat{a} < \hat{b}$ *in Eqs. (13) and (14) exists.* □

Lemma 2.2 states the credible interval parameter estimation for θ_t based on Bayes decision theory. However, this is not a prediction problem for x_{t+1} but the parameter estimation problem. The credible interval prediction problem based on Bayes decision theory is formulated in the next section.

3 CREDIBLE INTERVAL PREDICTION BASED ON BAYES DECISION THEORY

This section formulates the credible interval prediction problem based on Bayes decision theory (Weiss and Blackwell, 1961; Berger, 1985). Defining the loss, the risk, and the Bayes risk functions, the Bayes optimal credible interval prediction $\left[\hat{a}^*, \hat{b}^*\right]$ for x_{t+1} is obtained as the minimizer of the Bayes risk function $BR(a,b)$.

Definition 3.1. *Loss Function for Credible Interval Prediction*

$$L_3(a,b,x_{t+1}) = \begin{cases} c_1(x_{t+1}-b)+c_3(b-a), & \text{if } b \le x_{t+1}; \\ c_3(b-a), & \text{if } a \le x_{t+1} \le b; \\ c_2(a-x_{t+1})+c_3(b-a), & \text{if } x_{t+1} \le a, \end{cases} \tag{22}$$

where $\frac{c_3}{c_1}+\frac{c_3}{c_2}<1$. □

Definition 3.2. *Risk Function*

$$R(a,b,\theta_{t+1}) = \sum_{x_{t+1}=0}^{+\infty} L_3(a,b,x_{t+1})\, p\left(x_{t+1} \mid \theta_{t+1}\right). \tag{23}$$

□

Definition 3.3. *Bayes Risk Function*

$$BR(a,b) = \int_0^{+\infty} R(a,b,\theta_{t+1})\, p\left(\theta_{t+1} \mid \boldsymbol{x}^t\right) d\theta_{t+1}. \tag{24}$$

□

Definition 3.4. *Bayes Optimal Credible Interval Prediction*

$$\hat{a}^* = \arg\min_a BR(a,b), \tag{25}$$

$$\hat{b}^* = \arg\min_b BR(a,b). \tag{26}$$

□

Theorem 3.1. *Bayes Optimality of Credible Interval Prediction*

If the loss function in Eq. (22) is defined in the credible interval prediction problem, then the Bayes optimal solution $\left[\hat{a}^*, \hat{b}^*\right], \hat{a}^* < \hat{b}^*$ *in Eqs. (25) and (26) uniquely exists where* $\hat{a}^*, \hat{b}^*$ *satisfies,*

$$\int_0^{\hat{a}^*} p\left(x_{t+1} \mid \boldsymbol{x}^t\right) dx_{t+1} = \frac{c_3}{c_2}, \tag{27}$$

$$\int_{\hat{b}^*}^{+\infty} p\left(x_{t+1} \mid \boldsymbol{x}^t\right) dx_{t+1} = \frac{c_3}{c_1}. \tag{28}$$

□

Proof of Theorem 3.1.

Same proof as Lemma 2.2 (Winkler, 1972). In Lemma 2.2, the objective function is the posterior distribution of parameter $p\left(\theta_t \mid \boldsymbol{x}^t\right)$ since the credible interval parameter estimation problem is considered. Only difference is that the objective function in Theorem 3.1 is the predictive distribution $p\left(x_{t+1} \mid \boldsymbol{x}^t\right)$ for credible interval prediction problem. □

With both Definition 2.7 and Theorem 3.1, the direct relationship between $100(1-\lambda)\%$ credible interval for x_{t+1} in Eq. (10) and loss function in Eq. (22) is obtained. Table 1 shows some parameter examples of λ, c_1, c_2, and c_3.

Table 1: Parameter Examples for Credible Intervals and Loss Functions.

λ	c_1	c_2	c_3
0.01	200	200	1
0.05	40	40	1
0.10	20	20	1

Based on the Theorem 3.1, the following Bayes optimal credible interval prediction algorithm is proposed.

Algorithm 3.1. *Proposed Algorithm*

1. *Define the parameters* c_1, c_2, *and* c_3 *for loss function in Eq.(22).*
2. *Estimate hyper parameter k in Eq. (2) from training data.*
3. *Set* $t=1$ *and define the hyper parameters* α_1, β_1 *for the initial prior* $p\left(\theta_1 \mid \alpha_1, \beta_1\right)$ *in Eq. (3).*
4. *Update the posterior parameter distribution* $p\left(\theta_t \mid \alpha_t, \beta_t, \boldsymbol{x}^t\right)$ *under both prior* $p\left(\theta_t \mid \alpha_t, \beta_t\right)$ *and observed test data* $\boldsymbol{x}^t$ *in Eqs. (7) and (9).*
5. *Calculate the predictive distribution* $p\left(x_{t+1} \mid \boldsymbol{x}^t\right)$ *in Eq. (9).*
6. *Obtain the Bayes optimal credible interval* $\left[\hat{a}^*, \hat{b}^*\right]$ *from Eqs. (27) and (28).*
7. *If* $t < t_{max}$, *then set* $(t+1) \leftarrow t$, *the prior* $p(\theta_{t+1}) \leftarrow p\left(\theta_t \mid \alpha_t, \beta_t, \boldsymbol{x}^t\right)$, *and back to 4.*
8. *If* $t = t_{max}$, *then terminate the algorithm.*

□

4 NUMERICAL EXAMPLES

This section shows numerical examples to evaluate the performance of Algorithm 3.1. Subsection 4.1 explains both training and test data specifications. Training data was used to estimate hyper parameter

k in Eq. (2). For estimation of $\hat{k}$, the empirical Bayes approach with the approximate maximum likelihood estimation is considered. Its detail is explained in subsection 4.2. The test was used to the credible interval estimation. Defined parameters and evaluation basis is described in subsection 4.3. Finally, results are shown in subsection 4.4.

4.1 Web Traffic Data Specifications

Table 2 and 3 show the training and test data specifications. Both web traffic data were obtained by recording the http request arrival time stamps every 3 minutes at the web server in the late Mar. 2005.

Table 2: Training Data Specifications.

Items	Values
Date	Mar. 25, 2005
Total Request Arrivals	$11,527$
Time Interval	Every 3 minutes
Total Time Intervals	$t_{max} = 305$

Table 3: Test Data Specifications.

Items	Values
Date	Mar. 26, 2005
Total Request Arrivals	$6,382$
Time Interval	Every 3 minutes
Total Time Intervals	$t_{max} = 291$

4.2 Hyper Parameter Estimation with Empirical Bayes Method

Since a hyper parameter $0 < k \leq 1$ in Eq. (2) is assumed to be known, it must be estimated for real data analysis. In this paper, the following maximum likelihood estimation with numerical approximation in terms of empirical Bayes method is considered to obtain $\hat{k}$.

$$\hat{k} = \arg\max_{k} L(k)\,, \tag{29}$$

$$\begin{aligned} L(k) & \\ &= p\left(x_1 \mid \theta_1\right) \prod_{i=2}^{t} p\left(x_i \mid \boldsymbol{x}^{i-1}, k\right) & (30) \\ &= p\left(x_1 \mid \theta_1\right) \\ &\quad \cdot \prod_{i=2}^{t} \left[\int_0^{+\infty} p\left(x_i \mid \boldsymbol{x}^{i-1}, \theta i, k\right) p\left(\theta_i \mid \boldsymbol{x}^{i-1}\right) d\theta_i \right] & (31) \\ &= p\left(x_1 \mid \theta_1\right) \\ &\quad \cdot \prod_{i=2}^{t} \left[\frac{(\beta_i)^{\alpha_i} \Gamma(\alpha_i + x_i)}{(\beta_i+1)^{\alpha_i+x_i} \Gamma(\alpha_i) x_i!} \Bigg|_{\beta_i = k^{i-1}\beta_1 + \sum_{j=1}^{i-1} k^{j-1}}^{\alpha_i = k^{i-1}\alpha_1 + \sum_{j=1}^{i-1} k^{i-j} x_j} \right]. & (32) \end{aligned}$$

In this data analysis, $t = 305$ from Table 2 was used in the $L(k)$ since the training data was applied to estimate k. The numerically estimated value $\hat{k}$ was shown in Table 6 of subsection 4.4.

4.3 Evaluations of Credible Interval Prediction

For evaluations of credible interval prediction, the proposed nonstationary and the conventional stationary Poisson distributions were considered. The Bayes optimal credible interval predictions were derived for both distributions.

Table 4 shows the initial prior distributions of $p\left(\theta_1 \mid \alpha_1, \beta_1\right)$ in Eq. (3). This initial condition corresponds to the following non-informative prior (Berger, 1985; Bernardo and Smith, 2000),

$$p(\theta_1) = \frac{1}{\theta_1}\,. \tag{33}$$

The following posterior calculation shows that the initial prior in Eq. (33) corresponds to α_1, β_1 in Table 4. By the Bayes theorem, the posterior $p\left(\theta_1 \mid x_1\right)$ becomes,

$$\begin{aligned} p\left(\theta_1 \mid x_1\right) &= \frac{p\left(x_1 \mid \theta_1\right) p(\theta_1)}{\int_0^{\infty} p\left(x_1 \mid \theta_1\right) p(\theta_1)} & (34) \\ &= \frac{\exp(-\theta_1)(\theta_1)^{x_1-1}}{\Gamma(x_1)}\,. & (35) \end{aligned}$$

Thus Eq. (35) shows that,
$\theta_1 | x_1 \sim Gamma(\alpha_1 = x_1, \beta_1 = 1)$.

Table 4: Defined Hyper Parameters for Prior distribution $p(\theta_1)$.

α_1	β_1
x_1	1

For the loss function in Eq. (22), Table 4 shows the defined parameters c_1, c_2, and c_3. For the proposed prediction of $\left[\hat{a}^*, \hat{b}^*\right]$, 95% credible interval is considered. This means that $\hat{a}^*$ and $\hat{b}^*$ correspond to the lower 2.5 percentile and the upper 2.5 percentile (or lower 97.5 percentile) points of the predictive distribution $p\left(x_{t+1} \mid \boldsymbol{x}^t\right)$, respectively.

Table 5: Defined Parameters for Credible Intervals and Loss Functions.

λ	c_1	c_2	c_3
0.05	40	40	1

4.4 Results

Table 6 shows the estimated hyper parameter $\hat{k}$ from training data. Based on this $\hat{k}$, the proposed algorithm 3.1 is applied to obtain the Bayes optimal credible interval prediction. Figure 1 shows its result. In Figure 1, the horizontal and vertical axes are the index of time interval $1 \leq t \leq 291$ and the number of request arrival, respectively. Furthermore, the orange bar is real request arrival x_t, the blue solid line is the proposed upper value of the credible interval prediction $\hat{b}^*$, the red dotted line is the proposed lower value of $\hat{a}^*$. The green solid and cyan dotted lines mean the upper and lower limits of the credible interval predictions from stationary Poisson distribution.

Table 6: Hyper Parameter Estimation from Training Data.

Item	Value
$\hat{k}$	0.805

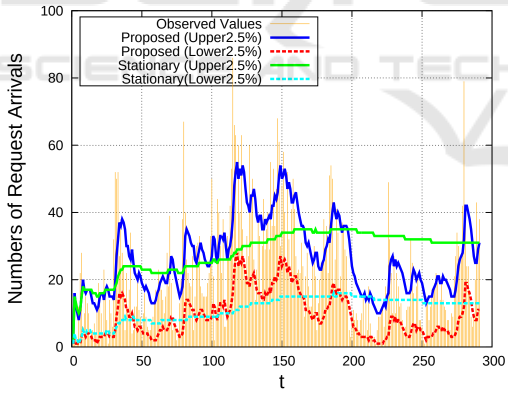

Figure 1: Credible Interval Prediction for Test Data.

Table 7 shows the number of time intervals satisfying $\hat{b}^* \geq x_t$ between the proposed nonstationary and conventional stationary Poisson distributions. Moreover, Table 8 shows the MSE between both models.

5 DISCUSSIONS

A hyper parameter k in the proposed nonstationary class in Eq. (2) and (5) generalizes a stationary Poisson distribution to one with a nonstationary parameter. If $k = 1$ in Eq. (2), $\theta_{t+1} = u_t\theta_t$ holds. In this case, $U_t \sim Beta[\alpha_t, 0]$ in Eq. (5). Since the second shape parameter in the Beta distribution becomes zero, the variance of u_t also becomes zero. This implies that the parameter θ_t in the Poisson distribution of x_t is stationary. However, if $0 < k < 1$ in Eq. (2), the parameter θ_t in the Poisson distribution is nonstationary.

In Eq. (9), β_t is expressed by the term $\sum_{i=1}^{t-1} k^i (x_i)^m$. This form is called the exponentially weighted moving average (EWMA) (Smith, 1979, p. 382),(Harvey, 1989, p. 350). The form is also observed in several versions of Simple Power Steady Model (Smith, 1979).

Theorem 3.1 implies that if the loss function in Eq. (22) is defined, then the unique Bayes optimal credible interval prediction $[\hat{a}^*, \hat{b}^*]$ is obtained. In this case, $\hat{b}^*$ is the upper $100(c_3/c_1)$ percentile point of the predictive distribution $p(x_{t+1} \mid \boldsymbol{x}^t)$. The $\hat{a}^*$ is the lower $100(c_3/c_2)$ percentile point of the $p(x_{t+1} \mid \boldsymbol{x}^t)$ likewise.

According to Table 6, the hyper parameter $\hat{k} = 0.805$ which is smaller than $k = 1$ is obtained. This implies that the training web traffic data can be estimated as the nonstationary Poisson distribution. Actually in Figure 1, the proposed prediction lines seem to flexibly follow the real traffic data well comparing to the stationary prediction lines.

Table 8 also shows that MSE of the proposed model is approximately 40% smaller than that of the stationary model. If $\lambda = 0.01, c_1 = c_2 = 200$, and $c_3 = 1$ from Table 1 are assumed to consider the upper 0.5% percentile $\hat{b}^*$, then this MSE is expected to be more smaller than that of the stationary model. Thus it can be concluded that the prediction performance of the proposed model can be relatively better than that of the stationary model. However, Table 7 shows that the coverage rate such that $\hat{b}^* \geq x_t$ of stationary Poisson distribution is greater than that of the proposed model. Therefore more precise predictions do not always help greater coverage rates.

Table 7: Number of Time Intervals with $\hat{b}^* \geq x_t$.

Items	Time Intervals with $\hat{b}^* \geq x_t$	Total Time Intervals	Coverage Rate
Stationary	212	291	72.9%
Proposed	208	291	71.5%

Table 8: Mean Squared Error between Upper 2.5 Percentile $\hat{b}^*$ and x_t for the Proposed and Stationary Models.

Items	MSE
Stationary	293.8
Proposed	185.2

6 CONCLUSION

This paper has proposed the credible interval prediction algorithm of a nonstationary Poisson distribution based on the Bayes decision theory. It is clarified that the Bayes optimal credible interval prediction can be uniquely obtained as the upper or lower percentile points of the predictive distribution under a certain loss function. Using real web traffic data, the performances of the proposed algorithm is evaluated. The upper limit of the credible interval prediction from the proposed nonstationary Poisson distribution has relatively smaller mean squared error by comparison with the stationary Poisson distribution.

The loss function defined in this paper has been restricted to the linear function. More generalized classes of loss functions can be considered and those would be future works.

REFERENCES

Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York.

Bernardo, J. M. and Smith, A. F. (2000). *Bayesian Theory*. John Wiley & Sons, Chichester.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Marsa, Malta.

Hogg, R. V., McKean, J. W., and Craig, A. T. (2013). *Introduction to Mathematical Statistics (Seventh Edition)*. Pearson Education, Boston.

Koizumi, D., Matsushima, T., and Hirasawa, S. (2009). Bayesian forecasting of www traffic on the time varying poisson model. In *Proceeding of The 2009 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA'09)*, volume II, pages 683–389, Las Vegas, NV, USA. CSREA Press.

Press, S. J. (2003). *Subjective and Objective Bayesian Statistics: Principles, Models, and Applications*. John Wiley & Sons, Hoboken.

Smith, J. Q. (1979). A generalization of the bayesian steady forecasting model. *Journal of the Royal Statistical Society - Series B*, 41:375–387.

Weiss, L. and Blackwell, D. (1961). *Statistical Decision Theory*. McGraw-Hill, New York.

Winkler, R. L. (1972). A decision-theoretic approach to interval estimation. *Journal of the American Statistical Association*, 67(337):187–191.

APPENDIX

A: Proof of Theorem 2.1

Note that time index t has been omitted for simplicity; for example, θ_t is written as θ, x_t is written as x, and so on. Suppose that data x are observed under the parameter θ following Eq. (2). Then, according to the Bayes theorem, the posterior distribution of the parameter $p\left(\theta \mid x\right)$ is as follows:

$$\begin{aligned} & p\left(\theta \mid x\right) \\ &= \frac{p\left(x \mid \theta\right) p\left(\theta \mid \alpha,\beta\right)}{\int_0^\infty p\left(x \mid \theta\right) p\left(\theta \mid \alpha,\beta\right) d\theta} \\ &= \frac{\frac{\beta^\alpha}{\Gamma(\alpha)x!} (\theta)^{\alpha+x-1} \exp\left[-(\beta+1)\theta\right]}{\frac{\beta^\alpha}{\Gamma(\alpha)x!} \int_0^\infty (\theta)^{\alpha+x-1} \exp\left[-(\beta+1)\theta\right] d\theta} \\ &= \frac{(\theta)^{\alpha+x-1} \exp\left[-(\beta+1)\theta\right]}{\int_0^\infty (\theta)^{\alpha+x-1} \exp\left[-(\beta+1)\theta\right] d\theta}. \qquad (36) \end{aligned}$$

Then the denominator of the right-hand side in Eq. (36) becomes,

$$\int_0^\infty (\theta)^{\alpha+x-1} \exp\left[-(\beta+1)\theta\right] d\theta = \frac{\Gamma(\alpha+x)}{(\beta+1)^{\alpha+x}}. \quad (37)$$

Note that Eq. (37) is obtained by applying the following property of the gamma function.

$$\frac{\Gamma(x)}{q^x} = \int_0^\infty y^{x-1} \exp(-qy)\, wt\,. \quad (38)$$

Substituting Eq. (37) in Eq. (36),

$$\begin{aligned} & p\left(\theta \mid x\right) \\ &= \frac{(\beta+1)^{\alpha+x}}{\Gamma(\alpha+x)} (\theta)^{\alpha+x-1} \exp\left[-(\beta+1)\theta\right]. (39) \end{aligned}$$

Eq. (39) shows that the posterior distribution of the parameter $p\left(\theta \mid x\right)$ also follows the gamma distribution with parameters $\alpha+x, \beta+1$, which is the same class of distribution as Eq. (3). This is the nature of the *conjugate family* (Bernardo and Smith, 2000) for the Poisson distribution.

Suppose the nonstationary transformation of the parameter θ in Eq. (2). Similar transformation of parameters for the Beta distribution is discussed (Hogg et al., 2013, pp. 162–163). According to Definition 2.6, the joint distribution $p(\theta,u)$ is the product of the probability distributions of θ in Eq. (3) and u in Eq. (5),

$$\begin{aligned} & p(\theta,u) \\ &= p(\theta)\, p(u) \\ &= \frac{(\beta)^\alpha}{\Gamma(k\alpha)\Gamma\left[(1-k)\alpha\right]} (u)^{k\alpha-1} (1-u)^{(1-k)\alpha-1} \\ & \quad \cdot (\theta)^{\alpha-1} \exp(-\beta\theta)\,. \qquad (40) \end{aligned}$$

Denote the two transformations as

$$\begin{cases} v = & \frac{\theta u}{k}; \\ & \\ w = & \frac{\theta(1-u)}{k}, \end{cases} \tag{41}$$

where $\theta > 0$, $0 < u < 1$, and $0 < k \leq 1$.

The inverse transformation of Eq. (41) becomes

$$\begin{cases} \theta = & k(v+w); \\ & \\ u = & \frac{v}{v+w}. \end{cases} \tag{42}$$

The Jacobian of Eq. (42) is

$$J = \begin{vmatrix} \frac{\partial\theta}{\partial v} & \frac{\partial\theta}{\partial w} \\ & \\ \frac{\partial u}{\partial v} & \frac{\partial u}{\partial w} \end{vmatrix} = \begin{vmatrix} k & k \\ & \\ \frac{w}{(v+w)^2} & -\frac{v}{(v+w)^2} \end{vmatrix} \tag{43}$$

$$= -\frac{k}{v+w} = -\frac{k^2}{\theta} \neq 0. \tag{44}$$

The transformed joint distribution $p(v,w)$ is obtained by substituting Eq. (42) for (40), and multiplying the right-hand side of Eq. (40) by the absolute value of Eq. (43):

$$\begin{aligned} & p(v,w) \\ = & \frac{(\beta)^{\alpha}}{\Gamma(k\alpha)\Gamma[(1-k)\alpha]}\left(\frac{v}{v+w}\right)^{k\alpha-1} \\ & \cdot\left(\frac{w}{v+w}\right)^{(1-k)\alpha-1}[k(v+w)]^{\alpha-1}\exp[-kb(v+w)] \\ = & \frac{(k\beta)^{\alpha}}{\Gamma(k\alpha)\Gamma[(1-k)\alpha]}(v)^{k\alpha-1}(w)^{(1-k)\alpha-1} \\ & \cdot\exp[-k\beta(v+w)]. \end{aligned} \tag{45}$$

Then, $p(v)$ is obtained by marginalizing Eq. (45),

$$\begin{aligned} p(v) &= \int_0^{\infty} p(v,w)\,dw \\ &= \frac{(k\beta)^{\alpha}(v)^{k\alpha-1}\exp(-k\beta v)}{\Gamma(k\alpha)\Gamma[(1-k)\alpha]} \\ & \quad \cdot\int_0^{\infty}(w)^{(1-k)\alpha-1}\exp(-k\beta w)\,dw \\ &= \frac{(k\beta)^{\alpha}(v)^{k\alpha-1}\exp(-k\beta v)}{\Gamma(k\alpha)\Gamma[(1-k)\alpha]}\frac{\Gamma[(1-k)\alpha]}{(k\beta)^{(1-k)\alpha}} \\ &= \frac{(k\beta)^{k\alpha}}{\Gamma(k\alpha)}(v)^{k\alpha-1}\exp(-k\beta v). \end{aligned} \tag{46}$$

Eq. (46) is obtained by applying the property of gamma function in Eq. (38).

According to Eq. (46), v follows the gamma distribution with parameters $k\alpha, k\beta$.

Considering two Eqs. (39) and (46), it has been proven that if the prior distribution of the scale parameter satisfies $\Theta \sim Gamma(\alpha,\beta)$, then its *transformed posterior distribution* satisfies

$$\Theta \mid x \sim Gamma[k(\alpha+x), k(\beta+1)]. \tag{47}$$

By adding the omitted time index t, the recursive relationships of the parameters of the gamma distribution can be formulated as,

$$\begin{cases} \alpha_{t+1} = & k(\alpha_t + x_t); \\ \beta_{t+1} = & k(\beta_t + 1). \end{cases} \tag{48}$$

Thus, for $t \geq 2$, the general α_t, β_t in terms of the initial α_1, β_1 can be written as

$$\begin{cases} \alpha_t = k^{t-1}\alpha_1 + \sum_{i=1}^{t-1} k^{t-i}x_i; \\ \beta_t = k^{t-1}\beta_1 + \sum_{i=1}^{t-1} k^{i-1}. \end{cases} \tag{49}$$

This completes the proof of Theorem 2.1. □

B: Proof of Theorem 2.2

From Eqs. (1) and (7), the predictive distribution under observation sequence $\boldsymbol{x}^t$ becomes,

$$\begin{aligned} & p(x_{t+1} \mid \boldsymbol{x}^t) \\ & = \int_0^{\infty} p(x_{t+1} \mid \theta_{t+1})\, p(\theta_{t+1} \mid \boldsymbol{x}^t)\, d\theta_{t+1} \end{aligned} \tag{50}$$

$$\begin{aligned} = & \int_0^{\infty}\left[\frac{\exp(-\theta_{t+1})}{x_{t+1}!}(\theta_{t+1})^{x_{t+1}}\right] \\ & \cdot\left[\frac{(\beta_{t+1})^{\alpha_{t+1}}}{\Gamma(\alpha_{t+1})}(\theta_{t+1})^{\alpha_{t+1}-1}\exp(-\beta_{t+1}\theta_{t+1})\right]d\theta_{t+1} \end{aligned} \tag{51}$$

$$\begin{aligned} = & \frac{(\beta_{t+1})^{\alpha_{t+1}}}{(x_{t+1})!\Gamma(\alpha_{t+1})} \\ & \cdot\int_0^{\infty}(\theta_{t+1})^{x_{t+1}+\alpha_{t+1}-1}[-(\beta_{t+1}+1)\theta_{t+1}]\,d\theta_{t+1} \end{aligned} \tag{52}$$

$$= \frac{(\beta_{t+1})^{\alpha_{t+1}}}{(x_{t+1})!\Gamma(\alpha_{t+1})}\frac{\Gamma(\alpha_{t+1}+x_{t+1})}{(\beta_{t+1}+1)}^{\alpha_{t+1}+x_{t+1}} \tag{53}$$

$$= \frac{\Gamma(\alpha_{t+1}+x_{t+1})}{(x_{t+1})!\Gamma(\alpha_{t+1})}\left(\frac{\beta_{t+1}}{\beta_{t+1}+1}\right)^{\alpha_{t+1}}\left(\frac{1}{\beta_{t+1}+1}\right)^{x_{t+1}}. \tag{54}$$

Note that Eq. (52) is obtained by applying the property of gamma function in Eq. (38).

This completes the proof of Theorem 2.2. □