On the Prediction of a Nonstationary Exponential Distribution Based on Bayes Decision Theory

Daiki Koizumi[a]
Otaru University of Commerce, 3–5–21, Midori, Otaru-city, Hokkaido, 045–8501, Japan

Keywords: Probability Model, Bayes Decision Theory, Nonstationary Exponential Distribution, Hierarchical Bayesian Model.

Abstract: A prediction problem with a nonstationary exponential distribution based on the Bayes decision theory was considered in this paper. The proposed predictive algorithm is based on both posterior and predictive distributions in a Bayesian context. The predictive estimator satisfies the Bayes optimality, which guarantees a minimum average error rate with a nonstationary probability model, a squared loss function, and a prior distribution of parameter. Finally, the predictive performance of the proposed algorithm was evaluated via comparison with the stationary exponential distribution using real meteorological data.

1 INTRODUCTION

The exponential distribution (Johnson et al., 1994; Bernardo and Smith, 2000) is a continuous probability distribution that has applications in various fields such as queuing theory (Kleinrock, 1975; Allen, 1990; Ross, 1997), reliability engineering (Gnedenko et al., 1969; Trivedi, 1982; Ross, 1997), and Bayesian statistics (Bernardo and Smith, 2000; Press, 2003). The stationary exponential distribution can be defined by both a nonnegative continuous random variable and a nonnegative parameter. The stationary exponential distribution has the so-called *memoryless* property, which leads to an independent distribution of service time in queuing theory (Allen, 1990, p. 123, 3.2.2), and a constant failure rate of lifetime distribution in reliability theory (Trivedi, 1982, pp. 122–123). Furthermore, the maximum likelihood estimator of the stationary exponential distribution can be obtained via simple arithmetic calculations (Trivedi, 1982, p. 482, Example 10.8); this implies that it is tractable for parameter estimation with real data.

Nevertheless, in the field of Bayesian statistics, parameter estimation or prediction problems with the Bayesian approach often become intractable problems. This is because these problems require integral calculations in the denominator of the Bayes theorem depending on a known prior distribution of parameter. However, if the specific distribution of parameter is assumed to be the prior, complex integral calculations can be avoided. In Bayesian statistics, this specific class of prior is called a *conjugate family* (Berger, 1985, pp. 130–132) (Bernardo and Smith, 2000, pp. 265–267). The gamma distribution is the natural conjugate prior of the stationary exponential distribution (Bernardo and Smith, 2000, p. 438).

The aforementioned results are limited within the stationary probability distributions. If nonstationary probability distributions are assumed, the Bayesian estimation problems become more difficult and more intractable. In such cases, there is no guarantee of the existence of a natural conjugate prior. In this regard, an interesting class of nonstationary probability models has been proposed, referred to as the *Simple Power Steady Model (SPSM)* (Smith, 1979). The SPSM is a time-series model with a specific class of nonstationary parameters. Under SPSM, they have shown certain illustrative probability distributions called *linear expanding families* in which natural conjugate priors exist (Smith, 1979). With regard to probability distributions in linear expanding families, the author proposed the same nonstationary parameter classes as in SPSM, approximated maximum likelihood estimation method of hyperparameters, and similar updating rules for the posterior distribution of parameters (Koizumi, 2020; Koizumi, 2021). Furthermore, a Bayesian problem in the Bayes decision theory was considered. Using this approach, the predictive estimator satisfies *Bayes optimality*, which guarantees a minimum average error rate for predictions. This approach has been applied to both a nonstationary Poisson distribution (Koizumi, 2020) and nonstationary Bernoulli distribution (Koizumi, 2021). The former concerns not only the Bayes optimal point prediction

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

Koizumi, D.
On the Prediction of a Nonstationary Exponential Distribution Based on Bayes Decision Theory.
DOI: 10.5220/0011632500003393
In *Proceedings of the 15th International Conference on Agents and Artificial Intelligence (ICAART 2023) - Volume 3*, pages 193-201
ISBN: 978-989-758-623-1; ISSN: 2184-433X

but also the Bayes optimal credible interval prediction (Koizumi, 2020). Both approaches demonstrate that the stationary natural conjugate priors can be generalized to the aforementioned nonstationary class of parameters, and tractable predictions are possible if nonstationary hyperparameters are known (Koizumi, 2020; Koizumi, 2021).

This paper presents the application of the aforementioned approach to a nonstationary exponential distribution. The proposed nonstationary class of parameter only contains single hyperparameter, which can be estimated using an approximate maximum likelihood estimation. If the hyperparameter is known, it can be proven that the posterior distribution of parameter can be obtained by simple arithmetic calculations. This property is crucial for obtaining the predictive distribution using the Bayes theorem. Furthermore, a Bayes optimal prediction algorithm is proposed. Finally, evaluation of the predictive performances of the proposed algorithms are via comparison with the results of the stationary exponential distribution using real data is detailed.

The rest of this paper is organized as follows. Section 2 provides the basic definitions of the nonstationary exponential distribution and two lemmas in terms of Bayesian statistics. Section 3 details the Bayes optimal prediction algorithm with respect to the Bayes decision theory. Section 4 presents numerical examples using real data. Section 5 presents a discussion on the results of this paper. Section 6 presents the conclusions drawn in this paper.

2 PRELIMINARIES

2.1 Hierarchical Bayesian Modeling with Nonstationary Exponential 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 $X_t \sim Exponential\left(\lambda_t\right)$, where $\lambda_t > 0$, is a nonstationary parameter. Thus, the probability density function of the nonstationary exponential distribution $p\left(x_t \mid \lambda_t\right)$ is defined as follows:

Definition 2.1. *Nonstationary Exponential Distribution*

$$p\left(x_t \mid \lambda_t\right) = \lambda_t \exp\left(-\lambda_t x_t\right), \tag{1}$$

where $\lambda_t > 0$. □

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

Definition 2.2. *Nonstationary Class of Parameter*

$$\lambda_{t+1} = \frac{u_t}{k}\lambda_t, \tag{2}$$

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 $\Lambda_t = \lambda_t$ is a continuous random variable from a Bayesian viewpoint. The prior $\Lambda_1 \sim Gamma\left(\alpha_1, \beta_1\right)$, where $\lambda_1 > 0$, $\alpha_1 > 0$, and $\beta_1 > 0$. This prior distribution is defined as follows:

Definition 2.3. *Prior Gamma Distribution for* λ_1

$$p\left(\lambda_1 \mid \alpha_1, \beta_1\right) = \frac{\left(\beta_1\right)^{\alpha_1}}{\Gamma\left(\alpha_1\right)}\left(\lambda_1\right)^{\alpha_1 - 1}\exp\left(-\beta_1\lambda_1\right), \tag{3}$$

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

Definition 2.4. *Gamma Function*

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

where $b \geq 0$. □

$\forall t$, $U_t \sim Beta\left[k\alpha_t, \left(1-k\right)\alpha_t\right]$, 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

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

□

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

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

$$p\left(\lambda_t, u_t \mid \alpha_t\right) = p\left(\lambda_t \mid \alpha_t\right)p\left(u_t \mid \alpha_t\right). \tag{6}$$

□

2.2 Lemmas for Posterior and Predictive Distributions

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

Lemma 2.1. *Posterior Distribution of* λ_t

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

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

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

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

□

Proof of Lemma 2.1.

See APPENDIX A. □

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

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

where $\alpha_{t+1}, \beta_{t+1}$ *are given as Eqs. (8).* □

Proof of Lemma 2.2.

See APPENDIX B. □

3 MAIN RESULTS

3.1 Basic Definitions

This subsection defines the loss function, the risk function, the Bayes risk function, and the Bayes optimal prediction based on Bayes decision theory (Berger, 1985; Bernardo and Smith, 2000). In this framework, the Bayes optimal prediction guarantees the minimum average error rate under the defined probability model, the loss function, and the prior distribution of parameter.

First of all, the following squared loss function is defined.

Definition 3.1. *Squared Loss Function*

$$L(\hat{x}_{t+1}, x_{t+1}) = (\hat{x}_{t+1} - x_{t+1})^2. \tag{10}$$

□

Secondly, the risk function, which is the expectation of the previous loss function with respect to the sampling distribution, is defined.

Definition 3.2. *Risk Function*

$$R(\hat{x}_{t+1}, \lambda_{t+1}) = \int_0^{+\infty} L(\hat{x}_{t+1}, x_{t+1})\, p\left(x_{t+1} \mid \lambda_{t+1}\right) dc_{t+1}, \tag{11}$$

where $p\left(x_{t+1} \mid \lambda_{t+1}\right)$ is from Definition 2.1. □

Thirdly, the Bayes risk function, which is the expectation of the previous risk function with respect to the posterior distribution of parameter, is defined.

Definition 3.3. *Bayes Risk Function*

$$BR(\hat{x}_{t+1}) = \int_0^{+\infty} R(\hat{x}_{t+1}, \lambda_{t+1})\, p\left(\lambda_{t+1} \mid \boldsymbol{x}^t\right) d\lambda_{t+1}, \tag{12}$$

where $p\left(\lambda_{t+1} \mid \boldsymbol{x}^t\right)$ is the posterior distribution of parameter which is described in Theorem 2.1. □

Finally, the Bayes optimal prediction, which guarantees the minimum average error rate, is defined.

Definition 3.4. *Bayes Optimal Prediction*

The Bayes optimal prediction $\hat{x}^*_{t+1}$ is obtained by,

$$\hat{x}^*_{t+1} = \arg\min_{\hat{x}_{t+1}} BR(\hat{x}_{t+1}). \tag{13}$$

□

3.2 Bayes Optimal Prediction

This subsection proves a Theorem which shows that the Bayes optimal prediction can be obtained by simple arithmetic calculations under the nonstationary exponential distribution and with both the squared loss function and known hyperparameter.

Theorem 3.1. *Bayes optimal Prediction*

If the squared loss function in Definition 3.1 is defined, then, the Bayes optimal prediction $\hat{x}^*_{t+1}$ *satisfies,*

$$\hat{x}^*_{t+1} = \frac{\beta_{t+1}}{\alpha_{t+1}}, \tag{14}$$

where $\alpha_{t+1}, \beta_{t+1}$ *are given as Eqs. (8).* □

Proof of Theorem 3.1.

For parameter estimation problem under the squared loss function, the posterior mean is the optimal (Berger, 1985, p. 161, Result 3 and Example 1). For the prediction problem, the predictive mean, i.e. the expectation of the Bayes predictive distribution is identically the optimal under the squared loss function defined in Definition 3.1.

$$\hat{x}^*_{t+1} = E\left[x_{t+1} \mid \boldsymbol{x}^t\right] \quad (15)$$

$$= \int_0^{+\infty} x_{t+1}\, p\left(x_{t+1} \mid \lambda_{t+1}\right) dx_{t+1} \quad (16)$$

$$= \frac{1}{\lambda_{t+1}} \quad (17)$$

$$= E\left[\lambda_{t+1}\right]^{-1} \quad (18)$$

$$= \frac{\beta_{t+1}}{\alpha_{t+1}}. \quad (19)$$

Note that Eq. (17) is derived because the expectation of the exponential distribution in Eq. (1) is $1/\lambda_t$. Since there exists the parameter distribution in Bayesian statistics, $1/\lambda_t$ equals to the inverse of expectation of posterior distribution of $p\left(\lambda_t \mid \alpha_t, \beta_t, \boldsymbol{x}^{t-1}\right)$ in Eq. (18). On the other hand, the posterior distribution of λ_t is gamma distribution according to Lemma 2.1. Moreover, its expectation as the gamma distribution becomes $E[\lambda_{t+1}] = \alpha_{t+1}/\beta_{t+1}$. Therefore, the Bayes optimal prediction $\hat{x}^*_{t+1} = \beta_{t+1}/\alpha_{t+1}$ as shown in Eq. (19). This completes the proof. □

3.3 Hyperparameter Estimation with Empirical Bayes Method

Since a hyperparameter $0 < k \leq 1$ in Eq. (2) is assumed to be known, it must be estimated in practice. In this paper, the following maximum likelihood estimation in terms of empirical Bayes method (Carlin and Louis, 2000) is considered.

Let $l(k)$ be a likelihood function of hyperparameter k and $\hat{k}$ be the maximum likelihood estimator. Then, those two functions are defined as,

$$\hat{k} = \arg\max_k l(k), \quad (20)$$

$$l(k) = p\left(x_1 \mid \lambda_1\right) p(\lambda_1) \prod_{i=2}^{t} p\left(x_i \mid \boldsymbol{x}^{i-1}, k\right) \quad (21)$$

$$= \prod_{i=1}^{t} \left[\frac{\Gamma(\alpha_i+1)(\beta_i)^{\alpha_i}}{\Gamma(\alpha_i)(\beta_i+x_i)^{\alpha_i+1}}\right], \quad (22)$$

where α_i, β_i satisfy Eqs. (8).

Eq. (22) can not be solved analytically and then the approximate numerical calculation method should be applied. The detail with real data is discussed in 4.4.1.

3.4 Proposed Predictive Algorithms

This subsection proposes the predictive algorithm which calculates the Bayes optimal prediction based on Theorem 3.1.

Algorithm 3.1. *Proposed Predictive Algorithm*

1. *Estimate hyperparameter k in Eq. (2) by Eq. (22) from training data.*
2. *Set $t = 1$ and define hyperparameters α_1, β_1 for the initial prior $p\left(\lambda_1 \mid \alpha_1, \beta_1\right)$ in Eq. (3).*
3. *Update the posterior distribution of parameter $p\left(\lambda_t \mid \alpha_t, \beta_t, \boldsymbol{x}^t\right)$ under both prior distribution of parameter $p\left(\lambda_t \mid \alpha_t, \beta_t\right)$ and observed test data $\boldsymbol{x}^t$ in Eqs. (7) and (8).*
4. *Calculate the predictive distribution $p\left(x_{t+1} \mid \boldsymbol{x}^t\right)$ in Eq. (9).*
5. *Obtain the Bayes optimal prediction $\hat{x}^*_{t+1}$ from Eq. (14).*
6. *If $t < t_{max}$, then update $(t+1) \leftarrow t$, the prior $p(\lambda_{t+1}) \leftarrow p\left(\lambda_t \mid \alpha_t, \beta_t, \boldsymbol{x}^t\right)$, and back to 3.*
7. *If $t = t_{max}$, then terminate the algorithm.*

□

4 NUMERICAL EXAMPLES

4.1 Conditions and Criteria for Evaluation

The performance of Algorithm 3.1 with real data is evaluated. For the comparison, two types of the Bayes optimal prediction $\hat{x}^*_{t+1}$ are considered. The first is from the proposed algorithm with nonstationary exponential distribution and the second is from a conventional algorithm with stationary exponential distribution. For the criteria for evaluations, the following cumulative squared error based on the squared loss function in Definition 3.1 is defined.

Definition 4.1. *Cumulative Squared Error*

$$\sum_{t=1}^{t_{max}} L(\hat{x}_t, x_t) = \sum_{t=1}^{t_{max}} (\hat{x}_t - x_t)^2. \quad (23)$$

□

Nextly, the following three points are explained: real data, initial prior distribution of parameter, and hyperparameter estimation. For real data, time series meteorological data is considered. It consists of training and test data as described in 4.2. Training data is applied to estimate the hyperparameter k in Definition 2.2. Test data is applied to evaluate the aforementioned two predictions. For the initial prior distribution of parameter, values of hyperparameter α_1, β_1 in Definition 2.3 must be known. This point is descried in 4.3. For hyperparameter estimation, the empirical Bayes approach already explained in 3.3 is considered with real data.

4.2 Data Specifications

Meteorological data is obtained as the daily average temperature Celsius in Tokyo from January 1, 2019 to December 31, 2020 (Japan Meteorological Agency, 2020). Table 1 and 2 show both training and test data specifications. Figure 1 shows their time series plots. In Figure 1, the red line shows time series of training data and the blue line shows that of test data.

Table 1: Training Data Specifications.

Items	Values
Monitoring Point:	Tokyo
From:	January 1, 2019
To:	December 31, 2019
Total Days (t_{max}):	365

Table 2: Test Data Specifications.

Items	Values
Monitoring Point:	Tokyo
From:	January 1, 2020
To:	December 31, 2020
Total Days (t_{max}):	366

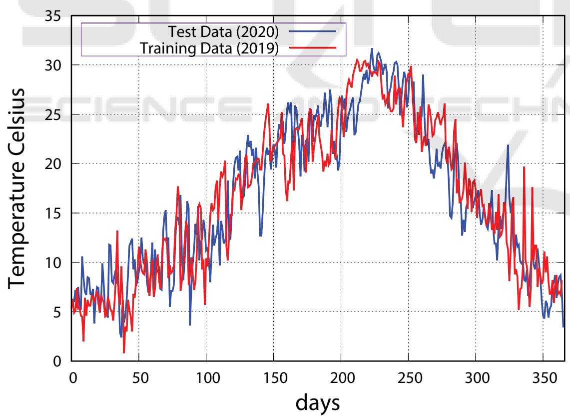

Figure 1: Time Series Plots of Training and Test Data.

4.3 Initial Prior Distribution of Parameter

According to both Definition 2.3 and Lemma 2.1, the class of the prior distribution of parameter is gamma distribution. If the non-informative prior (Berger, 1985; Bernardo and Smith, 2000) is considered under gamma prior, it is the exactly same condition as the author's previous paper considering the nonstationary Poisson distribution (Koizumi, 2020, p. 999, 4.3). Therefore, the detail explanation is omitted and the only hyperparameter settings in Definition 2.3 are shown in Table 3.

Table 3: Defined Hyperparameters for Prior distribution $p(\lambda_1)$.

Items	Values
α_1	x_1
β_1	1

4.4 Results

4.4.1 Hyperparameter Estimation

For the approximate maximum likelihood estimator of hyperparameter $\hat{k}$ in Eqs. (20) and (22), numerical calculation is executed with training data. Figure 3 and 4 show the plot for loglikelihood function $\log l(k)$ in Eq. (22). In Figure 2, the horizontal axis of $0 \leq k \leq 1$ shows the range of k which is divided into 1,000 subintervals and the vertical axis shows value of $\log l(k)$ in Eq. (22) with the logarithm base 10^3 which is required to avoid the numerical underflow. Figure 3 shows similar plot with enlarged horizontal axis of $0.92 \leq k \leq 1.00$. Finally, Table 4 shows the value of $\hat{k}$.

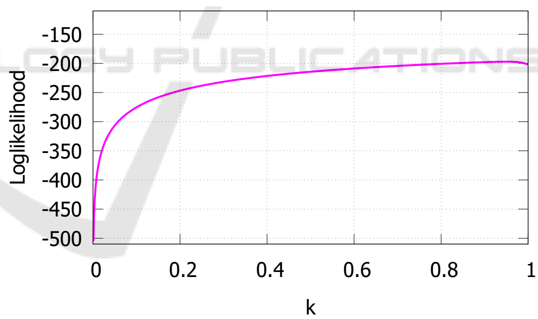

Figure 2: Loglikelihood Function $\log l(k)$ with the Logarithm Base 10^3 for $0 \leq k \leq 1$.

Table 4: Hyperparameter Estimation Result from Training Data.

Item	Value
$\hat{k}$	0.950

4.4.2 Bayes Optimal Prediction

Figure 4 shows time series plot of the observed data (orange bar), the prediction by the proposed model (blue line), and the prediction by the stationary model (red line) from test data. Table 5 shows the values

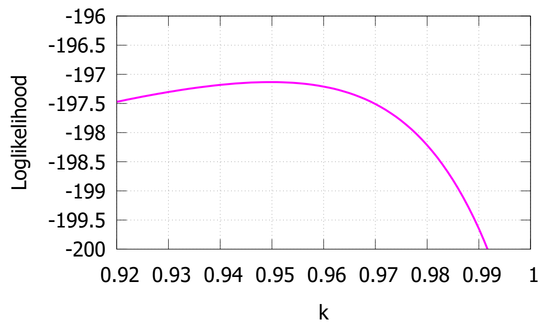

Figure 3: Loglikelihood Function $\log l(k)$ with the Logarithm Base 10^3 for $0.92 \leq k \leq 1.00$.

of cumulative squared error in Definition 4.1 for proposed and stationary models.

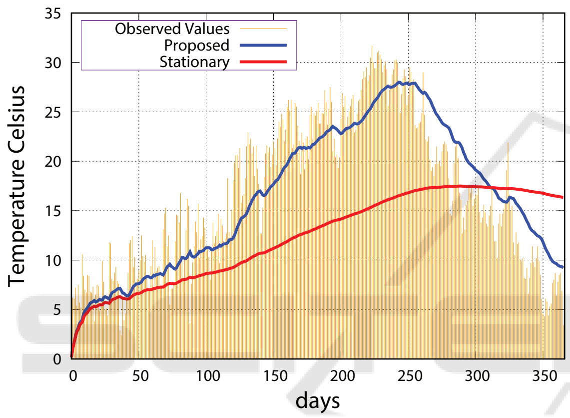

Figure 4: Predictions of the Proposed and Stationary Models for Test Data.

Table 5: Cumulative Losses for the Proposed and Stationary Models.

Items	Cumulative Squared Error
Stationary	62.4
Proposed	12.8

5 DISCUSSIONS

The hyperparameter k in Eq. (2) generalizes a stationary exponential distribution to a nonstationary one. If $k = 1$ in Eq. (2), then $\lambda_{t+1} = u_t \lambda_t$ holds. Consequently, $U_t \sim Beta[\alpha_t, 0]$ in Eq. (5). As the second shape parameter in the beta distribution becomes zero, the variance of u_t becomes zero as well. This means that the parameter λ_t in the exponential distribution of x_t is stationary. However, if $0 < k < 1$ in Eq. (2), then λ_t is nonstationary.

In Eqs. (8), β_t is expressed by the term $\sum_{i=1}^{t-1} k^{t-i} x_i$. This form is called the *Exponentially Weighted Moving Average (EWMA)* (Smith, 1979, p. 382),(Harvey, 1989, p. 350), which has also been observed in several versions of SPSMs (Smith, 1979; Koizumi, 2020; Koizumi, 2021).

Regarding the hyperparameter estimation with real data in 3.3, Figure 2 and 3 empirically show that the likelihood function $l(k)$ is upward convex in $0 \leq k \leq 1$. Even if a numerical calculation is required, the optimality of approximate maximum likelihood estimator $\hat{k}$ in Eq. (20) is partially guaranteed.

Regarding the Bayes optimal prediction, Figure 4 demonstrates that the prediction values of the proposed model follow a time series of test data more closely than those of the stationary model. In fact, Table 5 indicates that the value of the cumulative squared error of the proposed model is approximately twenty percent of that of the stationary model.

However, a more detailed analysis does not reveal that the proposed method is good prediction algorithm. The heights of the orange bars in Figure 4 basically increase until the 224th day. In fact, the highest daily average temperature is 31.7 [°C] on the 224th day, namely, August 11th in summer, 2020 in Tokyo. The blue line in Figure 4 also increases until 224th day. However, the 117 proposed predictions over 224 days, in other words 79.0%, are underestimated compared to the observed temperatures. On the other hand, after 225th day, the heights of the orange bars in Figure 4 start to decrease because the fall or winter seasons approach. The 113 proposed predictions over the remaining 142 days, namely, 79.6%, are overestimated compared to the observed temperatures. This result can be attributed to the fact that the loss function based on the Bayes decision theory is defined as the squared loss function stated in Eq. (10). The squared loss function is quadratic, and there is a significant predictive error if either overestimation or underestimation errors occur. In summary, the squared loss function yields the expectation of the predictive distribution with the Bayes optimal prediction, which often focuses on the *middle* range of observed data. If one is not satisfied with the above situation, another loss function should be defined.

6 CONCLUSION

This paper considered a specific class of nonstationary exponential distributions. We clarified that the Bayes optimal prediction governed by both the nonstationary distribution and squared loss function can be obtained through simple arithmetic calculations if the nonstationary hyperparameter is known. Using real meteorological data, the predictive performance of the proposed algorithm was proven superior to that of the stationary algorithm.

In the nonstationary hyperparameter estimation, the approximate maximum likelihood estimation is considered. We empirically observed that the likelihood function has an upward convexity regarding specific data. The general convexity should be proven, which is left for discussion in future research.

Moreover, the Bayes optimal prediction with loss functions other than the squared loss should be considered, which is another future research topic.

REFERENCES

Allen, A. O. (1990). *Probability, Statistics, and Queueing Theory with Computer Science Applications (Second Edition)*. Acaremic Press, San Diego.

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.

Carlin, B. and Louis, T. (2000). *Bayes and Empirical Bayes Methods for Data Analysis (Second Edition)*. Chapman & Hall, New York.

Gnedenko, B. V., Belyayew, Y. K., and Solovyev, A. D. (1969). *Mathematical Methods of Reliability Theory*. Acaremic Press, New York.

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.

Japan Meteorological Agency (2020). ClimatView (in Japanese). https://www.data.jma.go.jp/gmd/cpd/monitor/dailyview/. Browsing Date: Oct. 10, 2022.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions Volume 1 (Second Edition)*. Willey & Sons, New York.

Kleinrock, L. (1975). *Queueing Systems Volume I:Theory*. John Wiley & Sons, New York.

Koizumi, D. (2020). Credible interval prediction of a nonstationary poisson distribution based on bayes decision theory. In *Proceedings of the 12th International Conference on Agents and Artificial Intelligence - Volume 2: ICAART*, pages 995–1002. INSTICC, SciTePress.

Koizumi, D. (2021). On the prediction of a nonstationary bernoulli distribution based on bayes decision theory. In *Proceedings of the 13th International Conference on Agents and Artificial Intelligence - Volume 2: ICAART*, pages 957–965. INSTICC, SciTePress.

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

Ross, S. (1997). *Introduction to Probability Models (Sixth Edition)*. Academic Press, San Diego.

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

Trivedi, K. S. (1982). *Probability & Statistics with Reliability, Queuing and Computer Science Applications*. Prentice-Hall, Englewood Cliffs.

APPENDIX

A: Proof of Lemma 2.1

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

$$
\begin{aligned}
& p\left(\lambda \mid x\right) \\
&= \frac{p\left(x \mid \lambda\right) p\left(\lambda \mid \alpha, \beta\right)}{\int_0^{+\infty} p\left(x \mid \lambda\right) p\left(\lambda \mid \alpha, \beta\right) d\lambda} \\
&= \frac{\frac{\beta^\alpha}{\Gamma(\alpha)} (\lambda)^{\alpha-1} \exp\left[-(\beta+x)\lambda\right]}{\frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^{+\infty} (\lambda)^{\alpha-1} \exp\left[-(\beta+x)\lambda\right] d\lambda} \\
&= \frac{(\lambda)^{\alpha-1} \exp\left[-(\beta+x)\lambda\right]}{\int_0^{+\infty} (\lambda)^{\alpha-1} \exp\left[-(\beta+x)\lambda\right] d\lambda}. \qquad (24)
\end{aligned}
$$

Then the denominator of the right-hand side in Eq. (24) becomes,

$$
\int_0^{+\infty} (\lambda)^{\alpha-1} \exp\left[-(\beta+x)\lambda\right] d\lambda = \frac{\Gamma(\alpha+1)}{(\beta+x)^{\alpha+1}}. \qquad (25)
$$

Note that Eq. (25) is obtained by applying the following property of the gamma function.

$$
\frac{\Gamma(x)}{q^x} = \int_0^{+\infty} y^{x-1} \exp(-qy)\, dy. \qquad (26)
$$

Substituting Eq. (25) in Eq. (24),

$$
\begin{aligned}
& p\left(\lambda \mid x\right) \\
&= \frac{(\beta+x)^{\alpha+1}}{\Gamma(\alpha+1)} (\lambda)^{\alpha-1} \exp\left[-(\beta+x)\lambda\right]. \qquad (27)
\end{aligned}
$$

Eq. (27) shows that the posterior distribution of the parameter $p\left(\lambda \mid x\right)$ also follows the gamma distribution with parameters $\alpha+1, \beta+x$, which is the same class of distribution as Eq. (3). This is the nature of the *conjugate family* (Bernardo and Smith, 2000) for exponential distribution.

Suppose the nonstationary transformation of the parameter λ 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 $p(\lambda, u)$ is the product of the probability distributions of λ in Eq. (3) and u in Eq. (5),

$$\begin{aligned} p(\lambda, u) &= p(\lambda)\, p(u) \\ &= \frac{(\beta)^{\alpha}}{\Gamma(k\alpha)\,\Gamma[(1-k)\,\alpha]}\,(u)^{k\alpha-1} \\ &\quad \cdot (1-u)^{(1-k)\alpha-1}(\lambda)^{\alpha-1}\exp(-\beta\lambda)\,. \end{aligned} \tag{28}$$

Denote the two transformations as

$$\begin{cases} v = & \frac{\lambda u}{k}; \\ w = & \frac{\lambda(1-u)}{k}, \end{cases} \tag{29}$$

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

The inverse transformation of Eq. (29) becomes

$$\begin{cases} \lambda = & k(v+w); \\ u = & \frac{v}{v+w}. \end{cases} \tag{30}$$

The Jacobian of Eq. (30) is

$$J = \begin{vmatrix} \frac{\partial\lambda}{\partial v} & \frac{\partial\lambda}{\partial w} \\ \frac{\partial u}{\partial v} & \frac{\partial u}{\partial w} \end{vmatrix} = \begin{vmatrix} k & k \\ \frac{w}{(v+w)^2} & -\frac{v}{(v+w)^2} \end{vmatrix} \tag{31}$$

$$= -\frac{k}{v+w} = -\frac{k^2}{\lambda} \neq 0. \tag{32}$$

The transformed joint distribution $p(v,w)$ is obtained by substituting Eq. (30) for (28), and multiplying the right-hand side of Eq. (28) by the absolute value of Eq. (31):

$$\begin{aligned} & p(v,w) \\ &= \frac{(\beta)^{\alpha}}{\Gamma(k\alpha)\,\Gamma[(1-k)\,\alpha]}\left(\frac{v}{v+w}\right)^{k\alpha-1} \\ &\quad \cdot \left(\frac{w}{v+w}\right)^{(1-k)\alpha-1}[k(v+w)]^{\alpha-1} \\ &\quad \cdot \exp[-k\beta(v+w)] \cdot \left|-\frac{k}{v+w}\right| \\ &= \frac{(k\beta)^{\alpha}}{\Gamma(k\alpha)\,\Gamma[(1-k)\,\alpha]}(v)^{k\alpha-1}(w)^{(1-k)\alpha-1} \\ &\quad \cdot \exp[-k\beta(v+w)]\,. \end{aligned} \tag{33}$$

Then, $p(v)$ is obtained by marginalizing Eq. (33),

$$\begin{aligned} p(v) &= \int_0^{+\infty} p(v,w)\,dw \\ &= \frac{(k\beta)^{\alpha}(v)^{k\alpha-1}\exp(-k\beta v)}{\Gamma(k\alpha)\,\Gamma[(1-k)\,\alpha]} \\ &\quad \cdot \int_0^{+\infty}(w)^{(1-k)\alpha-1}\exp(-k\beta w)\,dw \\ &= \frac{(k\beta)^{\alpha}(v)^{k\alpha-1}\exp(-k\beta v)}{\Gamma(k\alpha)\,\Gamma[(1-k)\,\alpha]}\,\frac{\Gamma[(1-k)\,\alpha]}{(k\beta)^{(1-k)\alpha}} \\ &= \frac{(k\beta)^{k\alpha}}{\Gamma(k\alpha)}(v)^{k\alpha-1}\exp(-k\beta v)\,. \end{aligned} \tag{34}$$

Eq. (34) is obtained by applying the property of gamma function in Eq. (26).

According to Eq. (34), v follows the gamma distribution with parameters $k\alpha, k\beta$.

Considering two Eqs. (27) and (34), it has been proven that if the prior distribution of the scale parameter satisfies $\Lambda \sim Gamma(\alpha, \beta)$, then its *transformed posterior distribution* satisfies,

$$\Lambda \mid x \sim Gamma[k(\alpha+1), k(\beta+x)]\,. \tag{35}$$

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

$$\begin{cases} \alpha_{t+1} = & k(\alpha_t+1); \\ \beta_{t+1} = & k(\beta_t+x_t)\,. \end{cases} \tag{36}$$

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

$$\begin{cases} \alpha_t = k^{t-1}\alpha_1 + \sum_{i=1}^{t-1} k^i; \\ \beta_t = k^{t-1}\beta_1 + \sum_{i=1}^{t-1} k^{t-i}x_i. \end{cases} \tag{37}$$

This completes the proof of Lemma 2.1. □

B: Proof of Lemma 2.2

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

$$
\begin{aligned}
& p\left(x_{t+1} \mid \boldsymbol{x}^t\right) \\
& = \int_0^{+\infty} p\left(x_{t+1} \mid \lambda_{t+1}\right) p\left(\lambda_{t+1} \mid \boldsymbol{x}^t\right) d\lambda_{t+1} \quad (38) \\
& = \int_0^{+\infty} \left[\lambda_{t+1} \exp\left(-\lambda_{t+1} x_{t+1}\right)\right] \\
& \quad \cdot \left[\frac{\left(\beta_{t+1}\right)^{\alpha_{t+1}}}{\Gamma\left(\alpha_{t+1}\right)} \left(\lambda_{t+1}\right)^{\alpha_{t+1}-1} \exp\left(-\beta_{t+1}\lambda_{t+1}\right)\right] d\lambda_{t+1} \quad (39) \\
& = \frac{\left(\beta_{t+1}\right)^{\alpha_{t+1}}}{\Gamma\left(\alpha_{t+1}\right)} \\
& \quad \cdot \int_0^{+\infty} \left(\lambda_{t+1}\right)^{\alpha_{t+1}} \exp\left[-\left(\beta_{t+1} + x_{t+1}\right)\lambda_{t+1}\right] d\lambda_{t+1} \quad (40) \\
& = \frac{\left(\beta_{t+1}\right)^{\alpha_{t+1}}}{\Gamma\left(\alpha_{t+1}\right)} \frac{\Gamma\left(\alpha_{t+1}+1\right)}{\left(\beta_{t+1}+x_{t+1}\right)^{\alpha_{t+1}+1}} \quad (41) \\
& = \frac{\Gamma\left(\alpha_{t+1}+1\right)\left(\beta_{t+1}\right)^{\alpha_{t+1}}}{\Gamma\left(\alpha_{t+1}\right)\left(\beta_{t+1}+x_{t+1}\right)^{\alpha_{t+1}+1}} \quad (42)
\end{aligned}
$$

Note that Eq. (41) is obtained by applying the property of gamma function in Eq. (26).

This completes the proof of Lemma 2.2. □