The Economic Review (Otaru University of Commerce), vol. 71, No. 2 & 3, December, 2020.

Supplement to the paper "Credible Interval Prediction of a Nonstationary Poisson Distribution based on Bayes Decision Theory"

Daiki Koizumi*

Abstract

This article supplements the author's recent result [Koizumi, 2020] of numerical examples. In terms of Bayes decision theory [Berger, 1985], the Bayes optimal credible interval prediction is the minimizer of the Bayes risk function under the predefined loss function. Therefore, the predictive error should be measured based on not the mean squared error (MSE) but the mean of the above mentioned loss function. Furthermore, the effect of the prior distribution of the parameter (of the Poisson distribution) for the credible interval prediction is considered. Additionally, the errata of the original paper [Koizumi, 2020] are also presented.

* Department of Information and Management Science, Faculty of Commerce, Otaru University of Commerce

1 Corrected Numerical Example

1.1 Corrected Evaluations of Credible Interval Prediction

The author's recent result [Koizumi, 2020] used the mean squared error (MSE) to measure the predictive error, i.e.,

$$F_1\left(\hat{\boldsymbol{b}}^{*t}, \boldsymbol{x}^t\right) = \frac{1}{t}\sum_{i=1}^{t}\left(\hat{b}_i^* - x_i\right)^2 , \tag{1}$$

where $\hat{\boldsymbol{b}}^{*t} = \left(\hat{b}_1^*, \hat{b}_2^*, \ldots, \hat{b}_t^*\right)$ and $\boldsymbol{x}^t = (x_1, x_2, \ldots, x_t)$. Note that each $\hat{b}_i^*$ is the upper limit of the Bayes optimal credible interval at time i.

The above measurement is not consistent in terms of Bayes decision theory [Berger, 1985] since the specific loss function is predefined. This loss function should be used to evaluate predictive error. Thus the above F_1 should be replaced by the following function F_2,

$$F_2\left(\hat{\boldsymbol{a}}^{*t}, \hat{\boldsymbol{b}}^{*t}, \boldsymbol{x}^t\right) = \frac{1}{t}\sum_{i=1}^{t} L_3\left(\hat{a}_i^*, \hat{b}_i^*, x_i\right) , \tag{2}$$

where $\hat{\boldsymbol{a}}^{*t} = \left(\hat{a}_1^*, \hat{a}_2^*, \ldots, \hat{a}_t^*\right)$ and each $\hat{a}_i^*$ is the lower limit of the Bayes optimal credible interval at time i. Furthermore, $L_3\left(\hat{a}_i^*, \hat{b}_i^*, x_i\right)$ corresponds to $L_3\left(a, b, x_{t+1}\right)$ [Koizumi, 2020, Eq. (22), p. 998] and this $L_3\left(\hat{a}_i^*, \hat{b}_i^*, x_i\right)$ is defined as,

$$L_3\left(\hat{a}_i^*, \hat{b}_i^*, x_i\right) = \begin{cases} c_1\left(x_i - \hat{b}_i^*\right) + c_3\left(\hat{b}_i^* - \hat{a}_i^*\right), & \text{if } \hat{b}_i^* \leq x_i ; \\ c_3\left(\hat{b}_i^* - \hat{a}_i^*\right), & \text{if } \hat{a}_i^* \leq x_i \leq \hat{b}_i^* ; \\ c_2\left(\hat{a}_i^* - x_i\right) + c_3\left(\hat{b}_i^* - \hat{a}_i^*\right), & \text{if } x_i \leq \hat{a}_i^* , \end{cases} \tag{3}$$

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

1.2 Results

Table 1 shows corrected errors calculated by Eq. (2) and AICs (Akaike's Information Criterion [Akaike, 1973]) between stationary and proposed models.

Table 1: Corrected Predictive Error between $\left[\hat{a}_t^*, \hat{b}_t^*\right]$ and x_t

Items	Error	AIC
Stationary	233.7	$-3,839$
Proposed	138.0	$-4,009$

Table 2 shows detailed numbers of cases in Eq. (3). In Eq. (3), there are three cases depending on $\left(\hat{a}_i^*, \hat{b}_i^*, x_i\right)$. Each case is defined as the followings; (a) $\hat{b}_i^* \leq x_i$, (b) $\hat{a}_i^* < x_i < \hat{b}_i^*$, and (c)$x_i \leq \hat{a}_i^*$. Note that two equal signs are omitted in (b) in order to make the three cases exclusive.

Table 2: Detailed Numbers of Cases in Eq. (3)

Items	(a)	(b)	(c)	Subtotal
Stationary	83 (28.5%)	123 (42.3%)	85 (29.2%)	291 (100.0 %)
Proposed	94 (32.3%)	159 (54.5%)	38 (13.1%)	291 (100.0 %)

2 Supplementary Numerical Example

2.1 Effect Evaluation of Prior Distribution of Parameter

This section evaluates the effect of informative prior of parameter θ_t for credible interval prediction. For the initial prior distribution of parameter $p\left(\theta_1 \mid \alpha_1, \beta_1\right)$, only the non-informative prior is assumed [Koizumi, 2020, Eq. (33), p. 999]. In this article, the posterior distribution of parameter from training data is assumed as $p\left(\theta_1 \mid \alpha_1, \beta_1\right)$. Table 3 shows the obtained hyper parameters of α_1 and β_1 obtained from both the training data and Bayes theorem.

Table 3: Hyper Parameters for the Initial Prior distribution $p(\theta_1)$

α_1	β_1
74.06	5.13

2.2 Results

Table 4 shows the predictive error based on Eq. (2) with the initial prior in Table 3.

Table 4: Predictive Error between $\left[\hat{a}_t^*, \hat{b}_t^*\right]$ and x_t with Another Initial Prior

Items	Error
Stationary	234.2
Proposed	137.7

Table 5 shows detailed numbers of cases in Eq. (3) with the initial prior in Table 3.

Table 5: Detailed Numbers of Cases in Eq. (3) with Another Initial Prior

Items	(a)	(b)	(c)	Subtotal
Stationary	81 (27.8%)	124 (42.6%)	86 (29.6%)	291 (100.0 %)
Proposed	92 (31.6%)	158 (54.3%)	41 (14.1%)	291 (100.0 %)

3 Discussions

From Table 1, the predictive error of the proposed model is superior to that of stationary model with corrected loss function. The total error of the proposed model is about 41% smaller than that of stationary model. In terms of model selection, AIC value of the proposed model is smaller than that of stationary model. Since the smaller AIC value is more appropriate among candidate models, the proposed model is selected in terms of AIC. Note that the proposed model has only one extra parameter (hyper parameter k) comparing to stationary model.

Table 2 shows the detailed predictive performance between the proposed and stationary models. From Eq. (3), there are three cases ((a), (b), and (c)) to measure errors in credible interval prediction. According to Table 5 in the original paper [Koizumi, 2020, p. 1000], $c_1 = c_2 = 40$ and $c_3 = 1$ in Eq. (3). Note that case (a) has both c_1 and c_3, case (c) has both c_2 and c_3, and case (b) has only the smallest c_3 in Eq. (3). From Table 2, total number of case (b) in the proposed model is larger than that of stationary model and total number of case (c) in the proposed model is smaller than that of the stationary model. These points must contribute relatively smaller predictive error of the proposed model. On the other hand, total number of case (c) in the proposed model is larger than that of stationary model. This point means that the proposed model has relatively larger numbers of under estimations than that of stationary model. There exists a kind of the limitation in the proposed model with only one extra parameter.

From Table 4 and 5, almost no difference is observed comparing to Table 1 and 2 with both models, respectively. It can be concluded that no effect in prior settings was observed for author's evaluation with real data. Specifically for the proposed model, the estimated $\hat{k} = 0.805 < 1$ [Koizumi, 2020, Table 6, p. 1000] and the parameter of α_t in the Gamma distribution has the form of,

$$\alpha_t = k^{t-1}\alpha_1 + \sum_{i=1}^{t-1} k^{t-i} x_i \,, \tag{4}$$

in Eq. (8) [Koizumi, 2020, p. 996]. The Eq. (4) has the exponential form of k^{t-1} in the first term on the right hand side. Therefore it means that the larger the length of the observed data $\boldsymbol{x}^{t-1} = (x_1, x_2, \ldots, x_{t-1})$ becomes, the smaller the weight of α_1 from the initial prior $p(\theta_1)$. This point can be a kind of *memoryless* in the proposed model. On the other hand, it can be regarded as $k = 1$ in Eq. (4) for stationary model. In stationary model, the previous explanation with the exponential form of k^{t-1} can not be applied. However, note that the second term containing the sequence of x_i on the right hand side of Eq. (4). In this case, the larger the length of the observed data $\boldsymbol{x}^{t-1}$ becomes, the larger the weight of the second term in Eq. (4) becomes comparing to that of the first term from the initial prior. This point can be different kind of *memoryless* in stationary model. In summary, both the proposed and stationary models have lost the memory of the initial prior as the length of observed data increases. Therefore, non-informative prior can be simply applied for both models.

4 Errata

There are three typos in the paper.

1. Likelihood Function in Eq. (31) in Section 4.2 (p. 999)

   Incorrect:

$$
\begin{aligned}
&= p\left(x_1 \mid \theta_1\right) \\
&\quad \cdot \prod_{i=2}^{t} \left[\int_0^{+\infty} p\left(x_i \mid \boldsymbol{x}^{i-1}, \theta i, k\right) p\left(\theta_i \mid \boldsymbol{x}^{i-1}\right) d\theta_i \right]
\end{aligned}
\tag{31}
$$

   Correct:

$$
\begin{aligned}
&= p\left(x_1 \mid \theta_1\right) \\
&\quad \cdot \prod_{i=2}^{t} \left[\int_0^{+\infty} p\left(x_i \mid \boldsymbol{x}^{i-1}, \theta_i, k\right) p\left(\theta_i \mid \boldsymbol{x}^{i-1}\right) d\theta_i \right]
\end{aligned}
\tag{31}
$$

2. The number of end page in the fifth reference (p. 1001)

Incorrect: 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. CRESEA Press.

Correct: 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 683–689, Las Vegas, NV, USA. CRESEA Press.

3. Termination of integral term in Eq. (38) in Appendix A (p. 1001)

Incorrect:

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

Correct:

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

References

[Akaike, 1973] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Proceedings of the 2nd International Symposium on Information Theory*, pages 267–281. Akademiai Kiado.

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

[Koizumi, 2020] 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.