An analysis on determination of land area claim prediction for rice farmers insurance business in Indonesia using nonparametric bayesian method

I Maulidi$^1$, I Syahrini$^1$, R Oktavia$^1$, M Ihsan$^1$, R Emha$^1$

$^1$Department of Mathematics, Syiah Kuala University, Banda Aceh, Indonesia

*Correspondent author: ikhsanmaulidi@unsyiah.ac.id

Abstract. The Government of Indonesia has issued a Rice Farming Insurance program which is known as Asuransi Usaha Tani Padi (AUTP). It is a form of protection against crop failure losses faced by farmers by providing benefit insurance as the next planting capital. This study aims to estimate the Bayes premium parameters and Bayes premium estimates for the next period with 2 planting periods-claim data using a nonparametric approach. The results of this study were obtained by assuming a maximum area of crop failure or claims that occurred were 4 hectares for each group of farmers. Also, the analysis has been done by assuming the area of claimed land as a discrete random variable so that rounding of the data was required. Rounding data used in this study are those that are rounded to the nearest integers using the regular, flooring, and ceiling rounding method.

Keywords: Agricultural insurance, Bayesian analysis, Bayesian premium, Nonparametric model, Introduction

1. Introduction

Indonesia is known as one of the agrarian countries with the majority of the population earning a living in the agricultural sector which has an important role in the national economy. However, in reality, business in the agricultural sector is faced with the risk of uncertainty from various factors, such as flooding, drought, climate change, extreme weather, and crop pest organisms that cause high crop failure rates. Therefore, insurance is required to manage the crop failure risk [1,2]. In the USA, the agricultural insurance has been running for 85 years [3]. Meanwhile, in Indonesia, this program is a new financial product. The Ministry of Agriculture of the Republic of Indonesia has issued a Paddy Farming Insurance program or Asuransi Usaha Tani Padi (AUTP), which is expected to protect rice farmers from losses of crop failure by obtaining insurance compensation funds as the next planting capital [4].

It is important for policyholders that the premium paid to be fair and proportional to the estimated frequency of claims. Meanwhile, for the insurance company, the premium should produce a financial balance. In [5], research on estimating the number of claims and premiums on the AUTP program has been conducted. In this research, a method for estimating the number of premiums has been proposed using a parametric statistical approach by assuming that the data of rice yields follow the normal distribution. The performance of the method was evaluated using a Monte Carlo simulation based on the average value of the absolute error of the estimated premium to see the accuracy of the estimated results.
Another method that could be considered in determining premiums is the Bayesian method. Many types of research have been conducted using this method [6–10]. In [11] a model to determine agricultural insurance premium using the Bayesian method has been studied using a parametric approach. In [12] the Kriging Bayesian method has been studied using spatial and nonparametric approaches. The determination of premiums with the Bayesian approach has also been carried out in other research, some of which can be seen at [13–15].

The nonparametric Bayesian method for estimating claims has been discussed in [16–19] that inspired us to conduct a similar model for predicting claims based on AUTP previous claims data which is known as Bayesian premium [14,19,20]. We proposed a nonparametric Bayesian model to predict the claim land areas that would probably occur on agricultural insurance in Indonesia using simple and limited data. The claim data used is AUTP data in 2015, which only has two variables, namely the average area of the first-period claims and the average area of the second-period claims where the data are not distributed with a particular distribution. The nonparametric approach used in this study is appropriate for the data since the data are not fundamental to a particular distribution and the data size is small [21]. The claim periods are determined by the harvest periods of Rice in the particular areas that have two planting periods annually.

2. Research Method

The data used in this study are secondary data from the Aceh Food Crops Agency. The data used are data on participants of Rice Farming Insurance (AUTP) for the year of 2015 in four districts (Aceh Besar, Aceh Utara, Bireuen, and Pidie) and claim data (crop failure) of Rice Farming Insurance (AUTP) of October 2015 - March 2016 planting season and April 2016 - September 2016 planting season. The values of the data consisting of the average areas of Rice fields claimed to experience crop failures in the seasons after rounded up (using the ceiling).

The steps of the research work carried out in this study are as follows:

First, we determined a frequency table of claims per farmer group based on the average area of crop failure for each region (risk). Then, we determined a claim probability table. The claim probability table was built based on the claim frequency table that has been formed in the previous process by using the equation below.

\[
P(X = x_j | \Theta = \theta_k) = \frac{n(x_{j1}|\theta_k) + n(x_{j2}|\theta_k)}{2n(\theta_k)}, \quad j = 1,2,3,4,5, \quad k = 1,2,3,4.
\]  

where, \(n(x_{j1}|\theta_k)\) is the frequency of farmers that claim land area \(x_j\) in period 1 in the region \(\theta_k\), \(n(x_{j2}|\theta_k)\) is the frequency of farmers that claim land area \(x_j\) in period 2 in the region \(\theta_k\), and \(n(\theta_k)\) is
the frequency of farmers that make a claim in region \( \theta_k \), where \( \theta_1 \) stated Aceh Besar region, \( \theta_2 \) stated Aceh Utara (North Aceh) region, \( \theta_3 \) stated Bireuen region, and \( \theta_4 \) stated Pidie region. Then, we calculated the average claim area for each period. The period intended in this study is the rice planting season in 2015. Period 1 for the planting season of October 2015-March 2016 and period 2 for the planting season of April 2016-September 2016.

\[
\bar{x}_i = \frac{\sum_{j=1}^{5} \sum_{k=1}^{4} x_j n(x_j|\theta_k)}{\sum_{k=1}^{4} n(\theta_k)}, \quad i = 1,2. \tag{2}
\]

where, \( n(x_j|\theta_k) \) is the frequency of farmers that make a claim for land area \( x_j \) on the period I in region \( \theta_k \), while \( x_j \) is the claim land area of a \( j-1 \) hectare, \( n(\theta_k) \) is the frequency of farmers that make a claim in region \( \theta_k \). We then calculated the probability of claims event in the region of study

\[
P(\theta = \theta_k) = \frac{m(\theta_k)}{\sum_{i=1}^{4} m(\theta_i)}, \quad k = 1,2,3,4. \tag{3}
\]

where \( m(\theta_k) \) is the total of the land area of all farmers in region \( \theta_k \). The next step was calculating the probability of claims on two periods of the season,

\[
P(\bar{x}_1, \bar{x}_2) = \sum_{k=1}^{4} P(\bar{x}_1, \bar{x}_2|\theta_k)P(\theta = \theta_k). \tag{4}
\]

Then, we determined the probability of claims in the area of study on two periods of the season,

\[
P(\theta = \theta_k|\bar{x}_1, \bar{x}_2) = \frac{P(\theta_k, \bar{x}_1, \bar{x}_2)}{P(\bar{x}_1, \bar{x}_2)}, \quad k = 1,2,3,4. \tag{5}
\]

The estimate of the land area of claims in the area of study in the next period is given below,

\[
E[X|\theta = \theta_k] = \sum_{j=1}^{5} x_j P(x_j|\theta_k), \quad k = 1,2,3,4. \tag{6}
\]

Meanwhile, the estimate of the Bayesian premium as a land area of claims based on the previous data is given by the following equation,

\[
E[X|\bar{x}_1, \bar{x}_2] = \sum_{k=1}^{4} P(\theta_k)\bar{x}_1, \bar{x}_2]E(X|\theta_k). \tag{7}
\]

Since the random variables used in the study are assumed to be discrete, rounding to integers is required. We analyze this model with three types of rounding namely regular rounding, flooring, and ceiling. In this model, it is assumed that the maximum area of crop failure or claims for each farmer group is 4 hectares.

3. Results and Discussion

3.1. Table of Conditional Probability of Claims

From the data of claim, a frequency table and probability table for random claim variable \( X \) and the regional risk variable \( (\Theta) \) is presented in Table 1 below.

| Region(\( \Theta \)) | Period | Land area claim (hectares) | Total |
|----------------------|--------|-----------------------------|-------|
|                      |        | \( x_1 = 0 \) | \( x_2 = 1 \) | \( x_3 = 2 \) | \( x_4 = 3 \) | \( x_5 = 4 \) |       |
| \( \theta_1 \)      | 1      | 52              | 18       | 5         | 4         | 0       | 79     |
|                     | 2      | 68              | 4         | 2         | 2         | 3       |        |
| \( \theta_2 \)      | 1      | 31              | 11        | 8         | 0         | 6       | 56     |
|                     | 2      | 51              | 2         | 2         | 1         | 0       |        |
Based on Table 1, it is possible to calculate the probability of claim or crop failure occurring in an area of $x_1 = 0$ hectares, $x_2 = 1$ hectares, $x_3 = 2$ hectares, $x_4 = 3$ hectares, and $x_5 = 4$ hectares for each study area. Using equation (1), Table 2 is obtained as follows.

| Region ($\Theta$) | Period | Period | Period | Period | Total |
|-----------------|--------|--------|--------|--------|-------|
| $\theta_3$      | 1      | 16     | 5      | 1      | 1     | 26    |
|                 | 2      | 23     | 0      | 0      | 3     |       |
| $\theta_4$      | 1      | 36     | 4      | 7      | 4     | 2     | 53    |

Table 2. The Probability of Land Area Claims

The average area of land that is experiencing crop failure for each period can be calculated using equation (2). Because the data used in this study only have two planting seasons, there are two values for the average area of land claimed.

$$\bar{x}_i = \frac{\sum_{j=1}^{5} \sum_{k=1}^{4} x_j n(x_{ij} | \theta_k)}{\sum_{k=1}^{4} n(\theta_k)}$$

Hence, we have:

$$\bar{x}_1 = \frac{149}{214} = 0.69626$$

In the same way, we obtained $\bar{x}_2 = 0.29439$. Since the random variables used in the study are assumed to be discrete. There are three types of rounding rules to estimate Bayes premium, three cases can be made, suppose case 1 is using the regular rounding, where $[\bar{x}_1] = 1$, $[\bar{x}_2] = 0$, meanwhile case 2 is using the flooring rounding where $[\bar{x}_1] = [\bar{x}_2] = 0$ and case 3 is using the ceiling rounding where $[\bar{x}_1] = [\bar{x}_2] = 1$.

3.3. The Probability of Claim on Two Periods Season

In determining the chances of a claim occurring in the study area, the assumption is that the claims that occur are proportional to the area of land registered and each region has the same characteristics and treatment. The probability of a claim (crop failure) in the study area can be calculated using equation (3) so that it is obtained that:

$$P(\Theta = \theta_3) = \frac{m(\theta_3)}{\sum_{i=1}^{4} m(\theta_i)} = \frac{702}{2064} = 0.34030.$$
3.4. The Probability of Claim on Two Periods Season

The probability of crop failure occurring for 2 periods of the planting season if it is known that the probability of crop failure in the area to be studied (risk) can be calculated using equation (4) by using the value of the probability of crop failure occurring in the next section. Because $X_1 = \bar{x}_1$ and $X_2 = \bar{x}_2$ are two independent events, then equation (4) can be written as follows:

$$P(\bar{x}_1, \bar{x}_2) = \sum_{k=1}^{4} P[(\bar{x}_1, \bar{x}_2) | \theta_k] P(\theta = \theta_k) = \sum_{k=1}^{4} P(\bar{x}_1 | \theta_k)P(\bar{x}_2 | \theta_k) P(\Theta = \theta_k).$$

The unconditional probability of land area claims for periods 1 and 2 will be calculated for the three cases mentioned in section 3.2. Based on equation (4) the calculation of the chance of crop failure for case 1 is as follows:

$$P(\bar{x}_1 = 1, \bar{x}_2 = 0) =$$
$$P(\bar{x}_1 = 1 | \theta_1)P(\bar{x}_2 = 0 | \theta_1)P(\theta_1) +$$
$$P(\bar{x}_1 = 1 | \theta_2)P(\bar{x}_2 = 0 | \theta_2)P(\theta_2) +$$
$$P(\bar{x}_1 = 1 | \theta_3)P(\bar{x}_2 = 0 | \theta_3)P(\theta_3) +$$
$$P(\bar{x}_1 = 1 | \theta_4)P(\bar{x}_2 = 0 | \theta_4)P(\theta_4) = 0.08612.$$

In the same way, we calculate for case 2 ($\bar{x}_1 = \bar{x}_2 = 0$), and case 3 ($\bar{x}_1 = \bar{x}_2 = 1$). Therefore the results are obtained as shown in Table 3 as follows.

**Table 3. The Probability of CLAIM on Two Periods Season**

| Case | $\bar{x}_1$ | $\bar{x}_2$ | $P(\bar{x}_1, \bar{x}_2)$ |
|------|-------------|-------------|--------------------------|
| 1    | 1           | 0           | 0.08612                  |
| 2    | 0           | 0           | 0.55945                  |
| 3    | 1           | 1           | 0.01393                  |

Based on table 3, it is found that the greatest chance of claiming is 0.55945 which occurs in case 2, namely when the average claims that occur in period 1 and period 2 are 0 hectares.

3.5. The Probability of Claim in Region of Study on Two Periods Season

The probability of a claim (crop failure) in the study area ($\Theta = \theta_k$) if the areas of crop failure that occurred in the two previous planting season periods are known, can be determined using equation (5). The calculation of the probability of a claim (crop failure) $\theta_1$ in case 1 if the chances of crop failure occurred in the previous 2 planting season periods are known, with the average area of crop failure for period 1 is $\bar{x}_1 = 1$ and period 2 is $\bar{x}_2 = 0$ is done as follows:

$$P[\theta_1 | (\bar{x}_1 = 1, \bar{x}_2 = 0)] = \frac{P[\theta_1 \cap (\bar{x}_1 = 1, \bar{x}_2 = 0)]}{P(\bar{x}_1 = 1, \bar{x}_2 = 0)}$$
$$= \frac{P(\bar{x}_1 = 1 | \theta_1)P(\bar{x}_2 = 0 | \theta_1)P(\theta_1)}{P(\bar{x}_1 = 1, \bar{x}_2 = 0)}$$
$$= \frac{0.13924}{0.08612} = 0.4178.$$

For results of other regions in case 1, case 2, and case 3 can be seen in Table 4.
Table 4. The Probability of CLAIM in Region of Study on Two Periods Season in case 1, case 2, and case 3

|   | Case 1 | Case 2 | Case 3 |
|---|--------|--------|--------|
| θ | θ_k = θ_k(x_1 = 1, x_2 = 0) | θ_k = θ_k(x_1 = 0, x_2 = 0) | θ_k = θ_k(x_1 = 1, x_2 = 1) |
| θ_1 | 0.41786 | 0.35087 | 0.47378 |
| θ_2 | 0.45408 | 0.53603 | 0.44521 |
| θ_3 | 0.06779 | 0.08140 | 0.05374 |
| θ_4 | 0.06027 | 0.12680 | 0.02727 |

3.6. Prediction of The Land Area of Claim in the Next Period

The land area of a claim or the number of crop failures that are expected to occur in the next period for each study area (θ = θ_k) does not depend on the average value of the claimed areas. The results of the calculation of the claim area in the study area in the next period applicable to all cases and can be calculated using equation (6).

\[
E[X|θ] = \sum_{j=1}^{5} x_j P(x_j|θ) = x_1 P(x_1|θ_1) + x_2 P(x_2|θ_1) + x_3 P(x_3|θ_1) + x_4 P(x_4|θ_1) + x_5 P(x_5|θ_1)
\]

\[
= 0 + 0.13924 + 0.0886 + 0.11394 + 0.07596 = 0.41774 \text{ Hectare.}
\]

Then, using the same equation, the conditional expected value for each risk is presented in Table 5.

Table 5. The Land Area of Claim on the Next Period

|   | E[X|θ] (Hectare) |
|---|----------------|
| θ_1 | 0.41774 |
| θ_2 | 0.53572 |
| θ_3 | 0.61540 |
| θ_4 | 0.50946 |

From Table 5, it is shown that the smallest area of crop failure that will occur in the next period is located in Aceh Besar with an area of 0.41774 hectares, and the largest area of crop failure that will occur in the next period is located in Bireuen with the area of 0.61540 Hectare.

3.7. Bayesian Premium Estimation

Bayes premium is the number of claims or crop failures that are expected to occur in the next whole period of the study area (θ). Based on the estimated land area in the next period for case 1 can be calculated using equation (7) as follows:

\[
E[X|\bar{x}_1 = 1, \bar{x}_2 = 0] = \sum_{k=1}^{5} P[θ_k|\bar{x}_1, \bar{x}_2] E(X|θ_k)
\]

\[
= P[θ_1|\bar{x}_1, \bar{x}_2] E(X|θ_1) + P[θ_2|\bar{x}_1, \bar{x}_2] E(X|θ_2) + P[θ_3|\bar{x}_1, \bar{x}_2] E(X|θ_3) + P[θ_4|\bar{x}_1, \bar{x}_2] E(X|θ_4)
\]

\[
= (0.41786)(0.41774) + (0.45408)(0.53572) + (0.06779)(0.61540) + (0.06027)(0.50946)
\]

\[
= 0.49024 \text{ Hectare.}
\]

Then the Bayes premium is calculated for case 2 and case 3. Furthermore, the Bayes premium (hectare) that has been obtained is multiplied by the insurance costs stipulated by the insurance for one planting season of Rp 6,000,000/ha, so that the Bayes premium is obtained (Rupiah) with the calculation results in Table 6.

Table 6. Bayesian Premium

| Case | x_1 | x_2 | E[X|x_1, x_2] (Hectare) | E[X|x_1, x_2] (Rupiah) |
|------|-----|-----|------------------------|------------------------|
| 1    | 1   | 0   | 0.49024                | 2,941.434              |
| 2    | 0   | 0   | 0.54843                | 3,290.563              |
| 3    | 1   | 1   | 0.48339                | 2,900.338              |
Based on Table 6, it can be seen that the largest Bayesian premium is obtained using case 2, namely the rounding down model (floor). This model will certainly give benefit from the company side.

4. Conclusion and Suggestion
In this article, one method has been given to determine agricultural insurance premiums using a nonparametric approach. Although this approach still has drawbacks to the problem of accuracy compared to the parametric model, this method is very easy to implement. From the analysis conducted, the company can choose three cases of data discretization by rounding off. The type of rounding data that gives the largest and smallest premium values sequentially is floor rounding and ceiling rounding. Further researchers can select three cases of data discretization using rounding.

References
[1] Vávrová E 2009 The current demand for the insurability of liability for damage to the natural environment Agric. Econ. 55 33–9
[2] Rao K N 2010 Index based crop insurance Agric. Sci. Procedia 1 193–203
[3] Liu Y and Ker A 2020 Rating Crop Insurance Contracts with Nonparametric Bayesian Model Averaging J. Agric. Resour. Econ. 45 244–64
[4] Yanauri R, Aji J M M and Rondhi M 2019 Risk aversion level influence on farmer’s decision to participate in crop insurance: A review Agric. Econ. (Czech Republic) 65 481–9
[5] Mutaqin A K, Kudus A and Karyana Y 2015 Metode Parametrik Untuk Menghitung Premi Asuransi Usahatani Padi (AUTP) di Indonesia Prosiding SNaPP, pp 15–23
[6] Nickl R and Söhl J 2017 Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions Ann. Stat. 45 1664–93
[7] Phillips, L. D., & Wisniewski T K 2017 Bayesian Models for Computer-Aided Underwriting Author ( s ): Lawrence D . Philips and Thomas K . Wisniewski Source : Journal of the Royal Statistical Society . Series D ( The Statistician ), Vol . 32 , No . 1 / 2 , Proceedings of the 1982 I . O . S . Ann 32 252–63
[8] SIHOTANG H T, Panggabean E and Zebua H 2019 Sistem Pakar Mendiagnosa Penyakit Herpes Zoster Dengan Menggunakan Metode Teorema Bayes 3
[9] SIHOTANG H T 2019 Sistem Pakar Untuk Mendiagnosa Penyakit Pada Tanaman Jagung Dengan Metode Bayes 3
[10] SIHOTANG H T 2019 Perancangan Aplikasi Sistem Pakar Diagnosa Diabetes Dengan Metode Bayes 3 36–41
[11] Ozaki V A 2009 Pricing farm-level agricultural insurance: A Bayesian approach Empir. Econ. 36 231–42
[12] Park E, Wade Brorsen B and Harri A 2019 Using Bayesian kriging for spatial smoothing in crop insurance rating Am. J. Agric. Econ. 101 330–51
[13] Verrall R J 2004 A Bayesian Generalized Linear Model for the Bornhuetter-Ferguson Method of Claims Reserving North Am. Actuar. J. 8 67–89
[14] Migon H S and Moura F A S 2005 Hierarchical Bayesian collective risk model: An application to health insurance Insur. Math. Econ. 36 119–35
[15] Ker A P, Tolhurst T N and Liu Y 2016 Bayesian Estimation of Possibly Similar Yield Densities: Implications for Rating Crop Insurance Contracts Am. J. Agric. Econ. 98 360–82
[16] Johnson M, Griffiths T L and Goldwater S 2007 Adaptor grammars: A framework for specifying compositional nonparametric Bayesian models Adv. Neural Inf. Process. Syst. 641–8
[17] Panousis K P, Chatzis S and Theodoridis S 2019 Nonparametric Bayesian deep networks with local competition 36th Int. Conf. Mach. Learn. ICML 2019 2019-June 8772–83
[18] Xing W, Elhabian S, Kirby R M, Whitaker R T and Zhe S 2020 Infinite ShapeOdds:
Nonparametric Bayesian Models for Shape Representations

[19] Jeong H and Valdez E A 2020 Bayesian credibility premium with GB2 copulas Depend. Model. 8 157–71

[20] Ozaki V A and Silva R S 2009 Bayesian ratemaking procedure of crop insurance contracts with skewed distribution J. Appl. Stat. 36 443–52

[21] Hong L and Martin R 2017 A Flexible Bayesian Nonparametric Model for Predicting Future Insurance Claims North Am. Actuar. J. 21 228–41

Acknowledgments
The authors thank LPPM Syiah Kuala University for the financial support of this research through Hibah Penelitian Lektor PNPB 2020 with contract no. 166/UN11.2.1/PT.01.03/PNBP/2020. The authors also gratefully thank the reviewer for the comments and suggestions that have improved this article.