

# A study on the Bayesian nonparametric model for predicting group health claims

Muna Mauliza<sup>a</sup>, Jimin Hong<sup>1,a</sup>

<sup>a</sup>Department of Statistics and Actuarial Science, Soongsil University, Korea

---

## Abstract

The accurate forecasting of insurance claims is a critical component for insurers' risk management decisions. Hierarchical Bayesian parametric (BP) models can be used for health insurance claims forecasting, but they are unsatisfactory to describe the claims distribution. Therefore, Bayesian nonparametric (BNP) models can be a more suitable alternative to deal with the complex characteristics of the health insurance claims distribution, including heavy tails, skewness, and multimodality. In this study, we apply both a BP model and a BNP model to predict group health claims using simulated and real-world data for a private life insurer in Indonesia. The findings show that the BNP model outperforms the BP model in terms of claims prediction accuracy. Furthermore, our analysis highlights the flexibility and robustness of BNP models in handling diverse data structures in health insurance claims.

**Keywords:** risk management, health insurance claims, forecasting, Bayesian parametric model, Bayesian nonparametric model, simulated data, real-world data

---

## 1. Introduction

Health insurance is becoming increasingly popular as healthcare costs rise around the world. People buy health insurance to reduce their risk of personal bankruptcy due to medical expenses. One of the most popular types of private health insurance is group health insurance. People tend to enroll in group health insurance offered by their employers because it often provides coverage at a lower premium than individual health insurance.

Meanwhile, insurers are faced with the challenge of accurately predicting healthcare claims for risk management, liability assessment, and premium calculation. Therefore, many studies related to claims forecasting have been conducted, but there have also been criticisms of each methodology. For instance, a standard regression model can be applied to predict claims, yet the assumptions regarding linearity and independence are often inappropriate for real-world data.

More recently, Bayesian parametric (BP) models, which use prior distributions to account for the complexity of claims data, have been proposed as an alternative to standard regression models. Bayesian models can be used for claims prediction and premium calculation through credibility. A well-known credibility model in actuarial science is the Bühlmann model (2005). However, As stated by Huang and Meng (2020), the BP models require data to meet certain distributional assumptions, which are often not met in real-world data. Hong *et al.* (2017) also showed that the inflexibility in considering BP model can lead to significant biases in model misspecification, particularly when it

---

<sup>1</sup>Corresponding author: Department of Statistics and Actuarial Science, Soongsil University, 369 Sangdo-Ro, Dongjak-Gu, Seoul 06978, Korea. E-mail: [jmhong@ssu.ac.kr](mailto:jmhong@ssu.ac.kr)

comes to prediction. Therefore, this has led to the application of Bayesian nonparametric (BNP) models to accommodate the inflexibility of BP models in fitting nonstandard features of the distributions with respect to the parameters of claims data. Based on the BNP models, Zehnwirth (1979) provided a new credibility model and its application to actuarial problems. The BNP models are constructed from nonparametric priors that have extensive support over the space of distributions. For example, the Dirichlet process (DP) is a popular choice for the BNP model. The application of the BNP model using the DP was first studied by Ferguson (1973). Subsequently, Müller *et al.* (1996) developed the DP into an infinite mixture model by identifying the concept of a DPMM or Dirichlet process mixture model. Some literature such as Kottas (2006) and Cheng and Yuan (2013) studied the DPMMs for data with a non-negativity constraint. In particular, Kottas (2006) discussed a DPMM with a Weibull distribution, while Cheng and Yuan (2013) discussed a DPMM with a lognormal distribution.

Moreover, Fellingham *et al.* (2015) proposed a BNP model, namely a DPMM with a non-negativity constraint, and implemented the model to actuarial problems. They demonstrated that the BNP model is superior to the BP model in predicting group health claims. Hong and Martin (2017) used a DPMM of lognormal distributions for predicting automobile insurance claims. They applied the MCMC method to fit the model to data and presented a comparison of the performance of the DPMM model and the gamma kernel density estimate which also has a good fit. In addition, Richardson and Hartman (2018) considered covariate information in the BNP model to predict individual healthcare claims by classifications of illness and showed that the BNP model can be used for pricing. A recent study of Huang and Meng (2020) also applied the BNP model with covariates to automobile insurance claims.

Although the BNP model has been shown to outperform the BP model, the application of the BNP model to predict group health claims is still rare compared to automobile insurance claims. One of the contributions of this research is to show using real data that the BNP model is more effective than the BP model in predicting group health insurance claims.

There are also fewer applications of the BNP model to insurance data from developing countries compared to developed countries. Previous studies such as Fellingham *et al.* (2015) and Richardson and Hartman (2018) applied the BNP model to healthcare data from insurers in a developed country, namely the U.S. Hong and Martin (2017) also used insurance data from the U.S., but for automobile insurance data. They demonstrated that the BNP model outperformed the BP model in predicting insurance claims of developed countries. Furthermore, Huang and Meng (2020) applied the BNP model to automobile insurance data from developed and developing countries such as Australia, France, and mainland China.

Therefore, this study extends Fellingham *et al.* (2015) by applying the BNP model to group health insurance data from a private life insurer in Indonesia, which is a developing country. Particularly, our data spans 2018 and 2019, and exhibits some distinct features such as a higher percentage of zero claims and a larger number of insured per group, compared to the U.S. data.

Our findings are as follows. First, we estimate models by applying BNP and BP models to real data in 2018, and then validating the prediction using the corresponding data from 2019. Consistent with Fellingham *et al.* (2015), our results indicate that the BNP model outperforms the BP model in predicting group health claims of an insurer in Indonesia, showing a greater accuracy. Second, in the Bayesian models, it is necessary to obtain posterior distributions of the parameters before predicting claims, and the integration over the distributions of parameters given the data is required. Since the integrating is computationally challenging, we choose to simulate the posterior distributions using Markov chain Monte Carlo (MCMC) methods. We refer to Gilks *et al.* (1995) for a general review of the MCMC method and the algorithms used in both models are provided in Appendix.

The remainder is organized as follows. Sections 2 and 3 demonstrate two proposed Bayesian

models for predicting group health claims. A Bayesian parametric (BP) model will be explained first and then extended to a nonparametric one. In Section 4, we present two numerical examples using the simulated and real data from a private life insurer in Indonesia to compare the performance between two models. At last, the conclusions are summarized in Section 5.

## 2. Bayesian parametric model

For modeling group health claims using a Bayesian model, we first define a likelihood function and a prior distribution of the parameters associated with this function. As noted by Fellingham *et al.* (2015), we develop the likelihood function by considering the probability of zero claims and claims paid above zero which can vary for each group. Claims paid above zero (claim severity) typically have a positively skewed and fat-tailed distribution. To deal with these problems, we choose a mixture density function with a probability of zero claims in group  $j$ , denoted by  $\pi_j$ , as a zero mass and a gamma distribution with parameters  $(\gamma_j, \theta_j)$  as a distribution of claims paid above zero. The likelihood function can be defined as follows:

$$\prod_{j=1}^{G_P} \prod_{k=1}^{N_j} \left[ \pi_j I(x_{jk} = 0) + (1 - \pi_j) f(x_{jk} | \gamma_j, \theta_j) I(x_{jk} > 0) \right], \quad (2.1)$$

where  $j$  is an index of group,  $G_P$  denotes the number of groups,  $k$  is an index of member within a group,  $N_j$  is the number of participants in group  $j$ , and  $x_{jk}$  is the claims paid per day of exposure per insured (daily claims paid) in group  $j$ , and  $f(x_{jk} | \gamma_j, \theta_j)$  denotes a density function at  $x_{jk}$  of a gamma distribution with  $\gamma_j$  as shape parameter and  $\theta_j$  as scale parameter. In more detail,  $f(x_{jk} | \gamma_j, \theta_j)$  can be written as below:

$$f(x_{jk} | \gamma_j, \theta_j) = \frac{1}{\theta_j^{\gamma_j} \Gamma(\gamma_j)} x_{jk}^{\gamma_j} \exp(-x_{jk}/\theta_j)$$

for  $\gamma_j > 0$ ,  $\theta_j > 0$ , and  $j = 1, 2, \dots, G_P$ .

In the Bayesian models, the choice of prior distributions is crucial since it affects the accuracy of the predictions. More informative priors will result in more accurate predictions. As the first step in specifying the prior distributions in the BP model, we need to choose the random effects distribution for all parameters such as  $(\pi_j, \gamma_j, \theta_j)$ . In such model, a random effect is a set of unobserved variables that are assumed to be generated from a probability distribution. The distribution of the random effects is known as the random effects distribution. The random effects distribution is a key component of the Bayesian hierarchical models which represents the variability between groups in the data. The random effects distribution is usually assigned a prior distribution.

Hence, considering the non-negativity constraint of the parameters in the BP model, we choose the random effects distributions of  $(\pi_j, \gamma_j, \theta_j)$  as below. In particular, for  $j = 1, 2, \dots, G_P$ ,

$$\pi_j | \mu_\pi \sim \text{Beta}(\mu_\pi, \sigma_\pi^2),$$

$$\gamma_j | \beta \sim \text{Gamma}(b, \beta),$$

$$\theta_j | \delta \sim \text{Gamma}(d, \delta),$$

where  $\mu_\pi$ ,  $\sigma_\pi^2$ ,  $b$ ,  $\beta$ ,  $d$ , and  $\delta$  are the hyperparameters. In accordance with Forbes *et al.* (2011), Beta distribution has density function as follows:

$$f(\pi_j | \mu_\pi, \sigma_\pi^2) = \frac{1}{\text{Be}(a_1, a_2)} \pi_j^{a_1-1} (1 - \pi_j)^{a_2-1}$$

with  $0 < \pi_j < 1$ ,  $a_1 = (\mu_\pi^2 - \mu_\pi^3 - \mu_\pi \sigma_\pi^2) / \sigma_\pi^2$ ,  $a_2 = (\mu_\pi - 2\mu_\pi^2 + 3\mu_\pi^3 - \sigma_\pi^2 + \mu_\pi \sigma_\pi^2) / \sigma_\pi^2$ , and  $\text{Be}(a_1, a_2) = \int_0^1 u^{a_1-1} (1-u)^{a_2-1} du$  for  $a_1 > 0$  and  $a_2 > 0$ .

Next, we need to specify the hyperprior distributions for the hyperparameters. In the context of a BP model, a hyperprior distribution is a prior distribution placed on the hyperparameters of the random effects distribution. We assume that  $\mu_\pi$  follows a uniform distribution between 0 and 1, while  $\beta$  and  $\delta$  follow inverse gamma distributions with parameters  $(2, A_\beta)$  and  $(2, A_\delta)$ , respectively. Using 2 as the shape parameter for the distributions of  $\beta$  and  $\delta$  will allow them to have infinite prior variance. In addition, we select fixed values for hyperparameters  $\sigma_\pi^2$ ,  $b$ , and  $d$ . More information about choosing hyperparameter values for our data is provided in Section 4.

In summary, the posterior for full parameters of the BP model is proportional to:

$$\begin{aligned} & \prod_{j=1}^{G_p} \prod_{k=1}^{N_j} \left[ \pi_j I(x_{jk} = 0) + (1 - \pi_j) f(x_{jk} | \gamma_j, \theta_j) I(x_{jk} > 0) \right] \\ & \times \left[ \prod_{j=1}^{G_p} \frac{\beta^{-b}}{\Gamma(b)} \gamma_j^{b-1} \exp\left(\frac{-\gamma_j}{\beta}\right) \frac{\delta^{-d}}{\Gamma(d)} \theta_j^{d-1} \exp\left(\frac{-\theta_j}{\delta}\right) \right] \\ & \times \frac{1}{\text{Be}(a_1, a_2)} \pi_j^{a_1-1} (1 - \pi_j)^{a_2-1} \Big] p(\mu_\pi) p(\beta) p(\delta), \end{aligned}$$

where  $p(\mu_\pi)$ ,  $p(\beta)$ , and  $p(\delta)$  are the hyperpriors for  $\mu_\pi$ ,  $\beta$ , and  $\delta$ .

As is known, high-dimensional integration and unfeasible computations are required to analyze posterior distributions of the BP model. An alternative is to use the MCMC method to generate samples from the posterior distribution. We describe the MCMC method used for the BP and BNP models in Appendix and its implementation on our data in Section 4.

### 3. Bayesian nonparametric model

The random-effects distributions for the parameters  $(\pi_j, \gamma_j, \theta_j)$  chosen for the BP model may not be flexible enough to accommodate the distribution of the actual claims data. Moreover, it is very difficult to choose an informative prior distribution that can fit the data well. Therefore, we assume that a nonparametric prior distribution is used instead to capture all possible distributional shapes of the actual claims data. This can happen because the BNP model allows the shape of the random effects distribution to be determined by the data.

As pointed by MacEachern (2016), the Dirichlet process (DP) prior is one of the most frequently used priors in the Bayesian models. The DP prior consists of two parameters: A precision parameter, denoted by  $\alpha > 0$ , which controls how close the prior distribution to  $G_0$  and a parametric baseline distribution  $G_0$  which is same as the mean of the DP prior. A larger  $\alpha$  value indicates a closer DP distribution to  $G_0$ . According to the DP constructive definition stated by Sethuraman (1994), if a distribution follows the DP prior, denoted by  $G \sim \text{DP}(\alpha, G_0)$ , then with probability 1,

$$G = \sum_{j=1}^{\infty} \omega_j \delta_{v_j},$$

where  $\delta_y$  is a point mass distribution at  $y$ , the  $v_j$  are i.i.d. from  $G_0$ , and  $\omega_j$  are the weights that can be

obtained through a stick – breaking process as below:

$$\begin{aligned}\omega_1 &= \zeta_1, \\ \omega_j &= \zeta_j \prod_{l=1}^{j-1} (1 - \zeta_l),\end{aligned}$$

where  $\zeta_1, \zeta_2, \dots, \sim \text{Beta}(1, \alpha)$ . It shows that the distribution  $G \sim \text{DP}(\alpha, G_0)$  is a discrete probability distribution.

Next, we will define the DP priors in the BNP model. We assume that the probability of zero claims and the distribution of claims in a group are independent. Therefore, we choose to handle parameters  $\pi_j$  and  $(\gamma_j, \theta_j)$  separately by setting  $G_1$  as the random effects distribution for  $\pi_j$  and  $G_2$  as the random effects distribution for  $(\gamma_j, \theta_j)$ . In detail, here are the random effects distributions of the BNP model for  $j = 1, 2, \dots, G_P$ :

$$\begin{aligned}\pi_j &| G_1 \sim G_1, \\ (\gamma_j, \theta_j) &| G_2 \sim G_2, \\ G_1 &\sim \text{DP}(\alpha_1, G_{10}), \\ G_2 &\sim \text{DP}(\alpha_2, G_{20})\end{aligned}$$

assuming each parameter is independent. The parameters,  $\alpha_1, \alpha_2 > 0$ , are the precision parameters and  $G_{10}$  and  $G_{20}$  are the centering or parametric baseline distributions of the DP priors.

For the distributions of  $G_{10}$  and  $G_{20}$ , we use the same distributions as in the BP model. In particular,

$$\begin{aligned}G_{10} &\sim \text{Beta}(\mu_\pi, \sigma_\pi^2), \\ G_{20} &\sim \text{Gamma}(b, \beta) \times \text{Gamma}(d, \delta),\end{aligned}$$

where the hyperparameters are assumed to follow the same distributions as in the BP model.

Next, the marginalized version of model can be found by integrating  $G_1$  and  $G_2$  over their DP priors to get posterior inference. Then, we get the posterior for full parameters of BNP model which is proportional to the following expression:

$$\begin{aligned}p(\beta)p(\delta)p(\mu_\pi)p(\pi_1, \dots, \pi_{G_P} | \mu_\pi)p((\gamma_1, \theta_1), \dots, (\gamma_{G_P}, \theta_{G_P}) | \beta, \delta) \\ \times \left[ \prod_{j=1}^{G_P} \pi_j^{N_{j0}} (1 - \pi_j)^{N_j - N_{j0}} \right] \times \left[ \prod_{j=1}^{G_P} \prod_{\{k: x_{jk} > 0\}} f(x_{jk}; \gamma_j, \theta_j) \right],\end{aligned}$$

where  $N_{j0} = |\{k; x_{jk} = 0\}|$  and  $N_j - N_{j0} = |\{k; x_{jk} > 0\}|$ .

To get posterior distributions using MCMC method, we need to obtain the priors for each  $\pi_j$  and  $(\gamma_j, \theta_j)$  first. As explained by Blackwell and MacQueen (1973), the DP priors can be constructed using the Polya urn scheme. In particular, for  $j = 1, 2, \dots, G_P$ ,

$$p(\pi_j | \{\pi_i : i \neq j\}, \mu_\pi) = \frac{\alpha_1}{\alpha_1 + G_P - 1} g_{10}(\pi_j; \mu_\pi, \sigma_\pi^2) + \frac{1}{\alpha_1 + G_P - 1} \sum_{i=1}^{G_P-1} I_{\pi_i}(\pi_j), \quad (3.1)$$

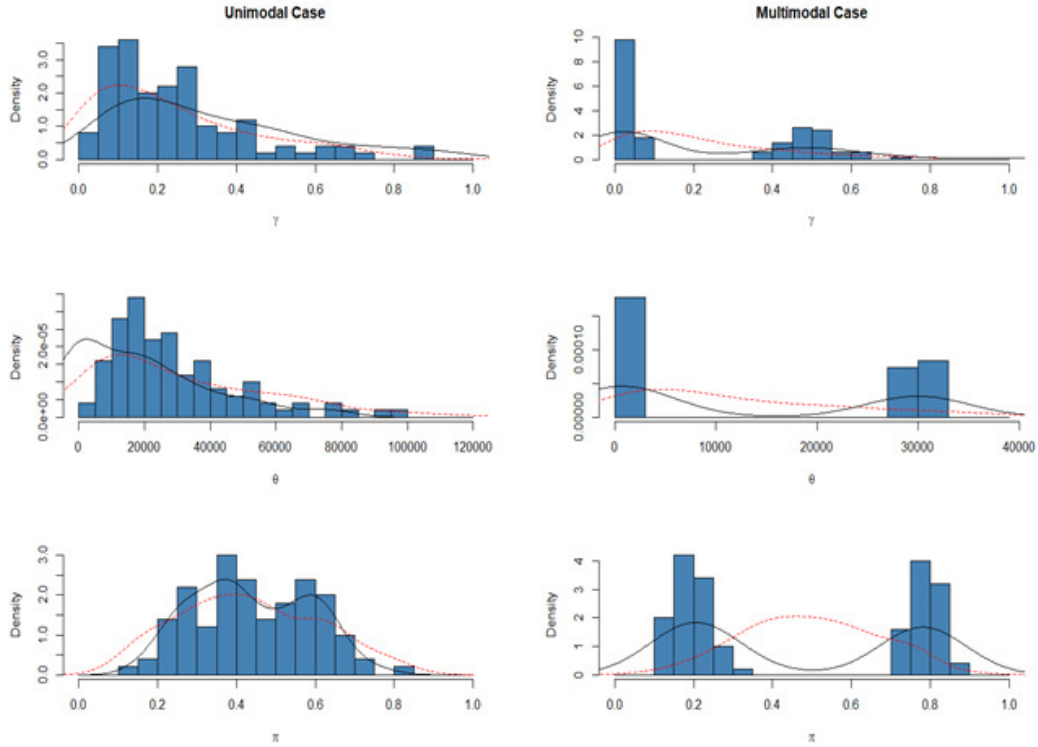


Figure 1: Simulated parameters of unimodal case and multimodal case.

and

$$p((\gamma_j, \theta_j) | \{(\gamma_i, \theta_i) : i \neq j\}, \beta, \delta) = \frac{\alpha_2}{\alpha_2 + G_P - 1} g_{20}((\gamma_j, \theta_j); \beta, \delta) + \frac{1}{\alpha_2 + G_P - 1} \sum_{i=1}^{G_P-1} I_{(\gamma_i, \theta_i)}(\gamma_j, \theta_j), \tag{3.2}$$

where  $g_{m0}$  is the density function corresponding to  $G_{m0}$  for  $m = 1, 2$ , and  $I(\cdot)$  is the indicator function.

The expressions in (3.1) and (3.2) are used as the proposal distributions to draw posterior samples of the parameters in the MCMC method. The parameters are generated from the centering distributions  $G_{m0}$  with probabilities  $\alpha_m/(\alpha_m + G_P - 1)$  or from the previous values of other parameters with probabilities  $(\alpha_m + G_P - 1)^{-1}$  for  $m = 1, 2$ . Then, posterior predictive samples can be generated by using the posterior samples of the parameters and according to likelihood function in (2.1).

## 4. Numerical data analysis

### 4.1. Simulated data

Two cases of data are generated to describe the possible distribution shapes of parameters in real data, particularly the unimodal case and the multimodal case. The random effects parameters of the unimodal case are drawn from unimodal distributions while the random effects parameters of the multimodal case are drawn from multimodal distributions.

Table 1: MSE of all groups in the simulated data

Model	Unimodal case	Multimodal case
BP	0.689	82.386
BNP	0.204	$9.254 \times 10^{-6}$

To obtain simulated data, we first generate the random effects parameters  $(\pi_j, \gamma_j, \theta_j)$  from the distributions mentioned below to reflect the real claims data used in Section 4.2. Then, simulated data are drawn for 100 groups and 100 observations in each group using these parameters and according to likelihood function (2.1). All draws are assumed to be independent.

The first case is the unimodal case. In here, the  $\pi_j$  are produced from a Beta distribution with parameters (4, 5), the  $\gamma_j$  from a Gamma distribution with parameters (2, 0.125), and the  $\theta_j$  from a Gamma distribution with parameters (2, 15000). While for the multimodal case, the  $\pi_j$  are generated from a mixture distribution of Beta(80, 20) and Beta(20, 80) with equal weight, the  $\gamma_j$  from a mixture distribution of Gamma(2, 100) and Gamma(50, 100) with equal weight, and the  $\theta_j$  from a mixture distribution of Gamma(1000, 1) and Gamma(30000, 1) with equal weight. After generating the parameters, the data are drawn using the likelihood function (2.1). After that, the data are analyzed under both the BP and BNP models.

For analyzing the BP model in both the unimodal case and the multimodal case, we use the following fixed values for hyperparameters, specifically  $\sigma_\pi^2 = 0.03$ ,  $b = d = 2$ ,  $A_\beta = 3$ ,  $A_\delta = 100$ . Even though, the sensitivity analyses shown by Fellingham *et al.* (2015) indicate that the other values of these parameters will actually result in the same posterior distributions. Furthermore, we remove the first 50,000 iterations (or use burn-in of 50,000 iterations) and keep every  $10^{th}$  draw from 100,000 posterior draws after burn-in.

For fitting the BNP model, we also apply the same values of hyperparameters for the centering distributions and choose  $\alpha_1 = \alpha_2 = 2$ . Then, we discard the first 10,000 iterations (or used burn-in of 10,000 iterations) since it takes longer time to get the result than in the BP model.

The plots of the simulated parameters and posterior densities under both BP and BNP models are given in Figure 1. The histograms plot the generated parameters, the dashed lines fit the BP model, and the solid lines fit the BNP model. We can see that both the BP and BNP models are able to capture the shapes of parameters quite closely in the unimodal case. While in the multimodal case, the BNP model is superior to the BP model in replicating the modes of parameters. The BP model will not replicate the modes of parameters unless we define it explicitly in the random effects distribution. On the other hand, defining multiple modes is not necessary under the BNP model since the DP priors can capture all distribution shapes of the parameters.

Next, following Albert (2009), we perform posterior predictive checks to check the convergence of the BP and BNP models. One way to perform the posterior predictive checks is to compare summary statistics or visualization of posterior predictive densities. First, we generate posterior predictive samples using the likelihood function (2.1) for all groups in the simulated data and compare the performance of both the BP and BNP models by comparing the mean square error (MSE) value.

The formula used for calculating MSE is as below:

$$\text{MSE} = \frac{1}{G_P \times N_j} \sum_{j=1}^{G_P} \sum_{k=1}^{N_j} (x_{jk} - \hat{x}_{jk})^2, \quad (4.1)$$

where  $x_{jk}$  is the claim amount drawn using generated random effects parameters  $(\pi_j, \gamma_j, \theta_j)$  while  $\hat{x}_{jk}$  is the claim amount drawn using simulated random effects parameters  $(\hat{\pi}_j, \hat{\gamma}_j, \hat{\theta}_j)$  under the BP or BNP

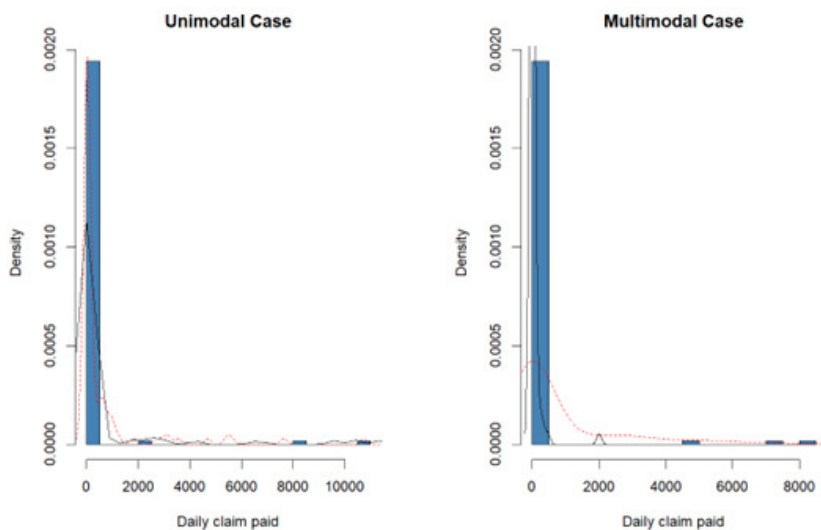


Figure 2: Cross-validated posterior predictive densities for simulated data.

Table 2: MSE of one group in the simulated data

Model	Unimodal case	Multimodal case
BP	4.680	0.042
BNP	0.419	0

models. The results of MSE values are shown in Table 1.

A better model will result in smaller MSE value. Table 1 displays that MSE values under the BNP model for both the unimodal case and the multimodal case are smaller than MSE values under the BP model. This indicates that the BNP model has greater accuracy in predicting claims of simulated data than the BP model.

Then, we choose one group from simulated data for both cases and draw posterior predictive densities as can be seen in Figure 2. The histograms plot the generated claims amount, the dashed lines fit the BP model, and the solid lines fit the BNP model. Obviously, we can observe that posterior predictive densities under the BNP model are closer to the real data than posterior predictive densities under the BP model.

Again, we also examine the MSE values as shown in Table 2. This table reveals the same results as before where the MSE values under the BP model are greater than the MSE values under the BNP model. Therefore, it can be concluded that the BNP model outperforms the similar BP model in predicting claims of simulated data.

## 4.2. Real data from a private life insurer in Indonesia

### 4.2.1. Summary of data

The actual data used is group health claims data from a private life insurer in Indonesia from policies in 2018 and 2019. All claim severities are presented in Indonesian currency, namely Indonesia Rupiah or IDR. As demonstrated by Fellingham *et al.* (2015), the proposed Bayesian models in this study are useful for modeling claims of medium-sized groups. For that reason, we filter the data by



Table 3: Descriptive statistics of real data from policies in 2018 and 2019

Data	Year	n insured	n group	Mean	Std. Dev.	Maximum	Skewness	Kurtosis	$\mu_\pi$
All	2018	9,338	112	3,224	13,440	463,333	13.47	284.93	67.18%
Groups	2019	8,843	47	4,951	21,101	728,832	15.72	379.38	56.09%
Company	2018	162	1	6,350	15,308	171,800	8.30	85.90	29.01%
F	2019	193		7,369	14,865	101,027	3.89	17.18	56.09%

excluding two large-sized groups with more than 10,000 insured per group. The types of insured in our data consist of employees and their dependents such as spouses and children. In more detail, Table 3 represents the descriptive statistics of real data in both 2018 and 2019 from all groups and one representative company that will be used in further analysis.

Table 3 provides some statistics such as the number of participants (insured), number of groups,  $\mu_\pi$  or the mean of proportion of zero claims per group as well as the mean, standard deviation, and maximum value of claims paid per number of days of policy in force per insured for all groups in the dataset. This table reveals that around 67% of all participants do not submit a claim in 2018. While in 2019, participants are more likely to submit claims compared to 2018 which can be seen in the smaller portion of zero claims. Other statistics like the mean, standard deviation, and maximum value of daily claims paid in 2019 are greater than in 2018. Besides, the number of groups in 2019 is less than in 2018 as well as the number of participants. In addition, skewness and kurtosis in Table 3 indicate that our data has right-skewed and heavy-tailed distributions in both years.

As previously mentioned, our data is from an insurer in a developing country. Since our data has some different characteristics compared to the data from an insurer in a developed country used by Fellingham *et al.* (2015), we want to check the efficacy of the BNP model over the BP model in modelling claims using our data. Some distinct characteristics are as follows. The proportion of zero claims from all groups in our data is around 60%, while in a developed country data is only around 30%. The average number of participants in the first year of our data and a developed country data is 83.3 and 8.3, respectively. Due to currency differences, our data has larger claims amount which will affect the distributions of the parameters. Also, we include employees, spouses, and children in our data, while they only include the employee plus one individual for each policy.

As mentioned before, our data is from an insurer in a developing country. Since our data has some different characteristics compared to the data from an insurer in a developed country used by Fellingham *et al.* (2015), we want to check the efficacy of the BNP model over the BP model in modelling claims using our data.

#### 4.2.2. Analysis

In the context of this study, we only analyze the renewal groups or the existing groups from the previous year. Also, we exclude some groups which have no claims and obtain 25 groups to be analyzed. For fitting both BP and BNP models, we use the same hyperparameters as simulated data, specifically  $\sigma_\pi^2 = 0.03$ ,  $b = d = 2$ ,  $A_\beta = A_\delta = 3$ . Then, we use burn-in of 50,000 iterations and keep every  $10^{\text{th}}$  draw from 100,000 posterior draws after burn-in for the BP model and use burn-in of 10,000 iterations for the BNP model. Also,  $\alpha_1 = \alpha_2 = 3$  are applied for the centering distributions of the BNP model.

Figure 3 below gives the plot of posterior distributions for  $\pi^*$ ,  $\gamma^*$ , and  $\theta^*$  for renewal groups where the solid and dashed lines fit the BNP and BP models, respectively. It is easy to see a multimodality behavior shown by the BNP model in the distributions of the parameters which is similar to the multimodal case described in Section 4.1. Under the BP model, it is almost impossible to find a

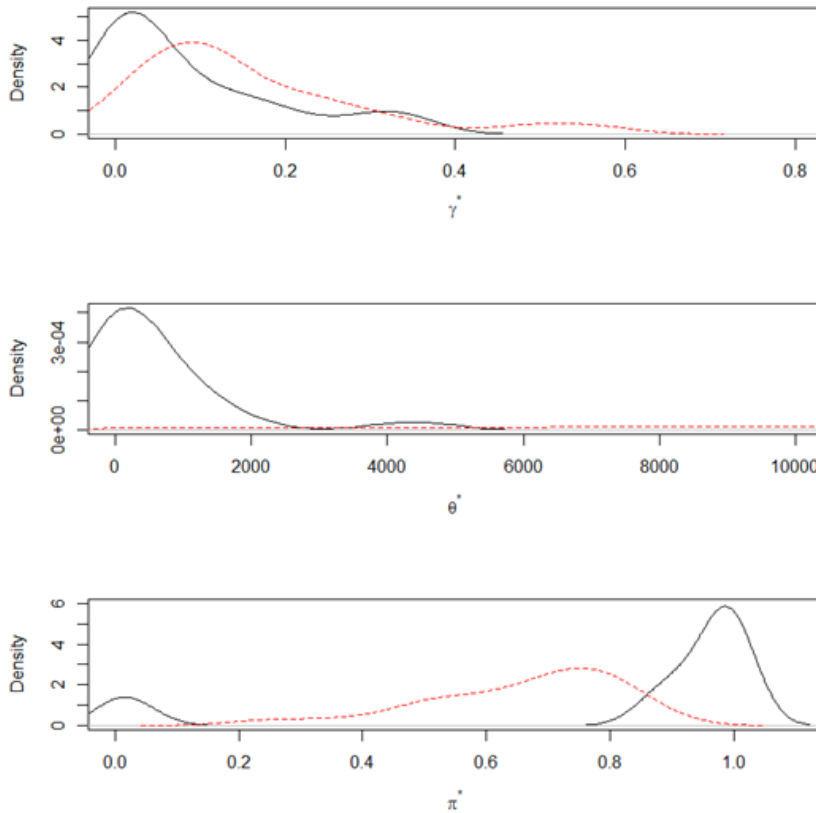


Figure 3: Posterior densities for real data.

multimodality behavior since it is in the distribution of the parameters not the data.

Furthermore, we choose one company called company F as the most representative group in 2018 and 2019 and predict its claims under the BP and BNP models. As displayed in Table 3, Company F has 162 participants in 2018 and 193 participants in 2019. We do not consider the presence of the same participants in both years. For the analysis, we obtain the posterior predictive densities of company F using the corresponding parameters  $(\pi_j, \gamma_j, \theta_j)$  and according to likelihood function in (2.1). The results are given in Figure 4 where the histograms here plot the daily claims paid of company F, the dashed lines fit the BP model, and the solid lines fit the BNP model.

Additionally, we compare both BP and BNP models by calculating the absolute value of mean difference between the actual and predicted claims of company F and MSE values of claims paid using the formula as in (4.1). We obtain the results as can be seen in Table 4.

In our results, under the BP model, it has an absolute value of mean difference of 3,430 and MSE of  $9.857 \times 10^5$ . While under the BNP model, it has a smaller absolute value of mean difference than the BP model at 1,358 and also a smaller MSE at  $1.928 \times 10^5$ . Obviously, this indicates that the predicted values under the BNP model are closer to real claims amount. This finding is in line with the results from simulated data and shows that the BNP model is better than the BP model in predicting group

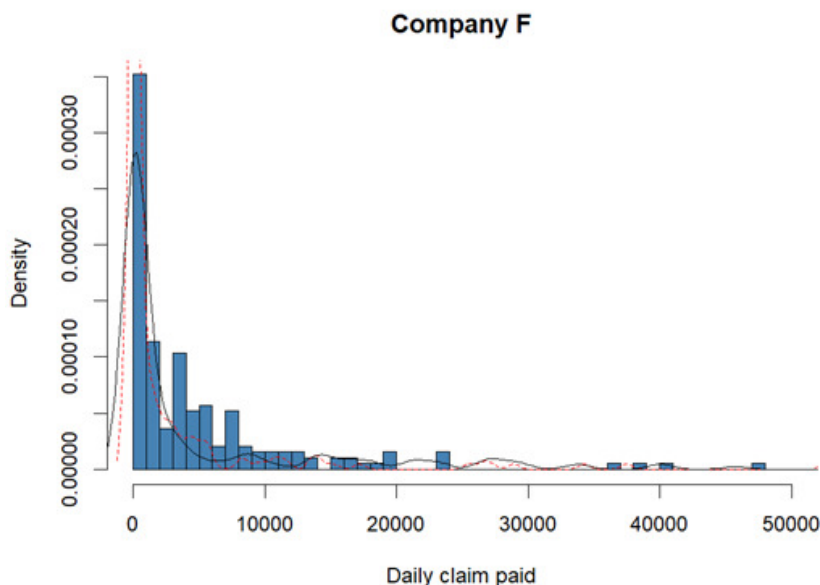


Figure 4: Cross-validated posterior predictive densities for real data.

Table 4: MSE of a representative group in real data

Model	$ E[X] - \mu $	MSE
BP	3,430	$9.857 \times 10^5$
BNP	1,358	$1.928 \times 10^5$

health claims of an insurer in Indonesia.

### 5. Conclusion

One of the most commonly used models to predict claims in the insurance industry is the BP model. However, the BP model is not flexible enough to accurately capture the complex distributions of health insurance claims. To overcome this limitation, the BNP model has been proposed, which employs a flexible prior distribution and offers a higher accuracy in predicting health insurance claims.

In the context of group health insurance, the BNP model proposed by Fellingham *et al.* (2015) is superior to the BP model in predicting claims in a developed country. The BNP model is capable of capturing the unique characteristics of health insurance claim distributions such as heavy tail and skewness, through the likelihood function, and can also accommodate the multimodality of the parameter distribution.

In this study, we implement this BNP model to predict group health insurance claims in Indonesia and compare its performance with a BP model. Our findings show that the BNP model is more suitable for describing non-standard forms of distributions, especially multi-modality behavior, in the posterior distributions of real data in a developing country. In contrast, the BP model cannot predict such behavior. Therefore, the use of the BNP model can improve risk management for insurers.

However, this study also has some limitations as follows. We only implement the Bayesian models for predicting actual claims for an insurer in a developing country for the renewal groups. To predict

claims for new business groups, different models need to be developed. Furthermore, since we only focus on the performance of the BP and BNP models, covariate information is omitted. Therefore, considering the possibility of covariates in all models can improve the results of this study in the future.

## Appendix A: MCMC method

As mentioned earlier, the posterior distributions for both BP and BNP models can be obtained by applying the MCMC method. Specifically, the Metropolis-Hastings (M-H) algorithm involves to solve multivariate distributions in the Bayesian models. The M-H algorithm produces a Markov chain based on a proposal distribution.

For instance, we want to produce samples from a posterior distribution  $p(\theta)$  which is proportional to  $g(\theta)$ , or denoted by  $p(\theta) \propto g(\theta)$ , then we need to select the initial value  $\theta_0$  and repeat the following steps for  $i = 1, \dots, m$  times.

1. Generate a candidate from a proposal distribution  $\theta^{new} \sim q(\theta^{new}|\theta_{i-1})$ , where the proposal distribution may rely on the value of  $\theta_{i-1}$  and can have any form.
2. Accept the candidate value or set  $\theta_i = \theta^{new}$  with probability:

$$\alpha = \min \left[ 1, \frac{g(\theta^{new})q(\theta_{i-1} | \theta^{new})}{g(\theta_{i-1})q(\theta^{new} | \theta_{i-1})} \right],$$

and  $\theta_i = \theta_{i-1}$  with probability  $1-\alpha$ .

We have to note that the distribution will be stationary to  $p(\cdot)$ . Discarded some samples or called burn-in samples is intended in the MCMC method to get its stationary distribution.

### MCMC algorithm for the BP model

Predicted data under the BP model can be obtained by drawing new parameters using the marginalized version of the model. We first do MCMC iterations to obtain the current values of hyperparameters. We use a normal proposal distribution with the current state as mean and a fixed variance to draw the new hyperparameters. Then, given the current hyperparameters, new parameters ( $\pi^*$ ,  $\gamma^*$ ,  $\theta^*$ ) are drawn from their random effects distributions. Next, predicted values are generated from function (2.1) using new parameters ( $\pi^*$ ,  $\gamma^*$ ,  $\theta^*$ ).

### MCMC algorithm for the BNP model

In the BNP model, we use density functions defined in (3.1) and (3.2) as the proposal distributions to draw new parameters of  $\pi^*$  and ( $\gamma^*$ ,  $\theta^*$ ) respectively for the renewal groups. Below are the details of MCMC algorithm to obtain  $\pi^*$  for each group  $j = 1, 2, \dots, G_p$ .

1. Let  $\pi_j^{old}$  be the current value of  $\pi_j^*$ .
2. Draw a candidate  $\pi_j^{new}$  from a proposal distribution defined in (3.1).
3. Accept the candidate value or  $\pi_j^* = \pi_j^{new}$  with probability:

$$q_1 = \min \left[ 1, \frac{(\pi_j^{new})^{N_{j0}} (1 - \pi_j^{new})^{N_j - N_{j0}}}{(\pi_j^{old})^{N_{j0}} (1 - \pi_j^{old})^{N_j - N_{j0}}} \right],$$

where  $N_{j0} = |k; x_{jk} = 0|$ ,  $N_j - N_{j0} = |k; x_{jk} > 0|$  and set  $\pi_j^* = \pi_j^{old}$  with probability  $1 - q_1$ .

4. Repeat the above steps a few times.

Next, we need to update the  $(\gamma_j, \theta_j)$  for each  $j = 1, 2, \dots, G_P$  following the steps below:

1. Let  $(\gamma_j^{old}, \theta_j^{old})$  be the current value of  $(\gamma_j^*, \theta_j^*)$ .
2. Draw a candidate  $(\gamma_j^{new}, \theta_j^{new})$  from a proposal distribution defined in (3.2).
3. Set  $(\gamma_j^*, \theta_j^*) = (\gamma_j^{new}, \theta_j^{new})$  with probability:

$$q_2 = \min \left[ 1, \frac{\prod_{\{k: x_{jk} > 0\}} f(x_{jk} | \gamma_j^{new}, \theta_j^{new})}{\prod_{\{k: x_{jk} > 0\}} f(x_{jk} | \gamma_j^{old}, \theta_j^{old})} \right],$$

and  $(\gamma_j^*, \theta_j^*) = (\gamma_j^{old}, \theta_j^{old})$  with probability  $1 - q_2$ .

4. Repeat the above steps a few times.

Then the predicted values for each renewal group are drawn from function (2.1) using new parameters  $(\pi^*, \gamma^*, \theta^*)$ .

### Appendix B: Burn-in length

As previously stated, burning of some samples is intended in the MCMC method. A longer burn-in may allow the chain to reach its stationary distribution and improve convergence. Therefore, to determine the appropriate burn-in length for the BNP model, we compare model performance using two burn-in lengths: 1,000 iterations and 10,000 iterations. The results of the comparison on simulated data can be seen in Table B.1. Meanwhile, the comparison for one group of real data is shown in Table B.2. Both tables indicate that using burn-in of 10,000 iterations is better than burn-in of 1,000 iterations for the BNP model. The large MSE value on real data is caused by the currency used for the claim amount.

Table B.1: MSE of BNP model for simulated data per burn-in length

Data	Unimodal case		Multimodal case	
	1,000 burn-in	10,000 burn-in	1,000 burn-in	10,000 burn-in
All groups	$5.295 \times 10^8$	0.204	$2.179 \times 10^8$	$9.254 \times 10^{-6}$
One group	$1.881 \times 10^8$	0.419	$0.839 \times 10^6$	0

Table B.2: MSE of BNP model for real data per burn-in length

Data	MSE	
	1,000 burn-in	10,000 burn-in
BNP	$4.655 \times 10^8$	$1.928 \times 10^5$

## References

- Albert J (2009). *Bayesian Computation with R*, Springer Science & Business Media, Berlin/Heidelberg.
- Blackwell D and MacQueen J (1973). Ferguson distributions via Polya Urn schemes, *The Annals of Statistics*, **1**, 353–355.
- Bühlmann H and Gisler A (2005). *A Course in Credibility Theory and Its Applications*, Springer, Berlin.
- Cheng N and Yuan T (2013). Nonparametric Bayesian lifetime data analysis using Dirichlet process lognormal mixture model, *Naval Research Logistics*, **60**, 208–221.
- Fellingham GW, Kottas A, and Hartman BM (2015). Bayesian nonparametric predictive modeling of group health claims, *Insurance: Mathematics and Economics*, **60**, 1–10.
- Ferguson TS (1973). A Bayesian analysis of some nonparametric problems, *The Annals of Statistics*, **1**, 209–230.
- Forbes C, Evans M, Hastings N, and Peacock B (2011). *Statistical Distributions*, John Wiley & Sons, New York.
- Gilks WR, Richardson S, and Spiegelhalter DJ (1995). *Markov Chain Monte Carlo in Practice*, Chapman and Hall/CRC, Florida.
- Hong L, Kuffner T, and Martin R (2016, December 10). On prediction of future insurance claims when the model is uncertain, *SSRN Electronic Journal*, Available from: <https://ssrn.com/abstract=2883574>
- Hong L and Martin R (2017). A flexible Bayesian nonparametric model for predicting future insurance claims, *North American Actuarial Journal*, **21**, 228–241.
- Huang Y and Meng S (2020). A Bayesian nonparametric model and its application in insurance loss prediction, *Insurance: Mathematics and Economics*, **93**, 84–94.
- Kottas A (2006). Nonparametric Bayesian survival analysis using mixtures of Weibull distributions, *Journal of Statistical Planning and Inference*, **136**, 578–596.
- MacEachern SN (2016). Nonparametric Bayesian methods: A gentle introduction and overview, *Communications for Statistical Applications and Methods*, **23**, 445–466.
- Müller P, Erkanli A, and West M (1996). Bayesian curve fitting using multivariate normal mixtures, *Biometrika*, **83**, 67–79.
- Richardson R and Hartman B (2018). Bayesian nonparametric regression models for modeling and predicting healthcare claims, *Insurance: Mathematics and Economics*, **83**, 1–8.
- Sethuraman J (1994). A constructive definition of Dirichlet priors, *Statistica Sinica*, **4**, 639–650.
- Zehnwirth B (1979). Credibility and the Dirichlet process, *Scandinavian Actuarial Journal*, **1**, 13–23.

Received September 7, 2023; Revised December 18, 2023; Accepted December 26, 2023