JKMA Jurnal Kajian Matematika dan Aplikasinya
VOLUME 2 NOMOR 2, July 2021
e-ISSN: 2722-7650. http://journal2.um.ac.id/index.php/jkma

PEMODELAN KLAIM ASURANSI MENGGUNAKAN PENDEKATAN BAYESIAN DAN MARKOV CHAIN MONTE CARLO

Azizah[1*]
[1] Jurusan Matematika Universitas Negeri Malang

Email : azizah.fmipa@um.ac.id
*Corresponding Author

Abstract

The determination of the correct prediction of claims frequency and claims severity is very important in the insurance business to determine the outstanding claims reserve which should be prepared by an insurance company. One approach which may be used to predict a future value is the Bayesian approach. This approach combines the sample and the prior information The information is used to construct the posterior distribution and to determine the estimate of the parameters. However, in this approach, integrations of functions with high dimensions are often encountered. In this Thesis, a Markov Chain Monte Carlo (MCMC) simulation is used using the Gibbs Sampling algorithm to solve the problem. The MCMC simulation uses ergodic chain property in Markov Chain. In Ergodic Markov Chain, a stationary distribution, which is the target distribution, is obtained. The MCMC simulation is applied in Hierarchical Poisson Model. The OpenBUGS software is used to carry out the tasks. The MCMC simulation in Hierarchical Poisson Model can predict the claims frequency

Keywords*: Markov Chain Monte Carlo, OpenBUGS, Hierarchical Poisson Model, Gibbs Sampling.*

Submitted: 9 April 2021; Revised: 3 Mei 2021; Accepted Publication: 11 Juni 2021;
Published Online: July 2021
DOI: 10.17977/um055v2i2p7-13

PENDAHULUAN

Salah satu bagian penting dalam ilmu aktuaria adalah prediksi ketidakpastian terhadap kejadian yang akan terjadi di masa mendatang. Prediksi ketidakpastian salah satunya diterapkan di dunia asuransi untuk mengestimasi banyak klaim maupun besar klaim (Scollnik, 2000). Hal ini, untuk selanjutnya digunakan sebagai dasar estimasi cadangan klaim suatu perusahaan. Estimasi cadangan klaim sangat penting karena sudah menjadi tanggung jawab perusahaan untuk memastikan kemampuan perusahaan membayar klaim – klaim yang diajukan oleh pemegang polis.

Proses prediksi banyak klaim maupun besar klaim dapat dilakukan dengan model peluang tertentu. Terdapat dua pendekatam dalam menentukan model peluang, yaitu pendekatan frekuentis dan pendekatan Bayesian (Cowless, 2013). Pada pendekatan bayesian, terdapat distribusi prediktif yang dapatdigunakan untuk memprediksi kejadian di masa yang akan datang. Untuk menentukan distribusi prediktif ini, terdapat tiga komponen penting yaitu distribusi prior, distribusi model atau distribusi *likelihood* dan distribusi posteior. Pendekatan bayesian menggunakan informasi saat ini yang diperoleh dari pengambilan sampel dan informasi masa lalu dari parameter (distribusi prior). Penggabungan ini menghasilkan distribusi posterior yang digunakan untuk menentukan distribusi prediktif.

Pendekatan Bayesian seringkali menjadi kompleks karena fungsi peluang memiliki dimensi yang tinggi atau memiliki banyak parameter dan seluruh parameter ini dianggap peubah acak (Klugman, 2004). Selain itu, fungsi yang dibangun terkadang memiliki bentuk yang jarang dikenal seperti uniform, gamma, eksponensial, dsb. Pada artikel ini, digunakan simulasi Markov Chain Monte Carlo (MCMC) untuk menentukan nilai integral dari suatu fungsi berdimensi tinggi. Metode ini menggabungkan konsep rantai Markov dan *Monte Carlo Integration*. Simulasi dilakukan pada *Hierarchical Poisson Model* untuk memprediksi banyak klaim pada bisnis asuransi umum.

Simulasi diawali dengan membangun suatu rantai markov ergodik dan distribusi posteriornya sebagai distribusi stasioner. Dengan kata lain, sampel yang dihasilkan dari sejumlah iterasi akan konvergen pada distribusi stationer yang diinginkan (Geman, 1992). Konsep ini memanfaatkan sifat apabila dilakukan suatu penarikan sampel dengan ukuran yang besar, barisan sampel yang diperoleh akan mengikuti distribusi tertentu. Sedangkan nilai integral dari fungsi yang diinginkan dapat dihitung menggunakan *Monte Carlo Integration.* Algoritma yang digunakan dalam simulasi MCMC ini adalah algoritma Gibbs Sampling dengan bantuan perangkat lunak OpenBUGS (*Open Bayesian Inference Using Gibbs Sampling*). OpenBUGS adalah perangkat lunak yang secara khusus digunakan untuk inferensi Bayesian menggunakan algoritma Gibbs Sampling.

METODE

Monte Carlo Integration

Metode Monte Carlo adalah salah satu metode numerik yang menggunakan penarikan sampel untuk mengapproksimasi suatu nilai. Integral Monte Carlo menggunakan proses ini untuk mengapproksimasi nilai integralnya (Klugman, 2004). Metode Monte Carlo dapat digunakan untuk menghitung nilai integral suatu fungsi peluang yang sulit untuk dipecahkan secara analitik. Dalam bisnis asuransi, fungsi peluang digunakan pada model kerugian. Model ini dapat diformulasikan dengan kerangka kerja statistik dan penyelesaianya dapat diestimasi secara empirik menggunakan metode Monte Carlo.

Misalkan, akan dihitung nilai integral dari fungsi berikut

$$I = \int_a^b h(\theta) f(\theta) d(\theta) \tag{1}$$

$f(\theta)$ adalah fungsi kepadatan peluang dari peubah acak θ dan $h(\theta)$ adalah suatu fungsi dari peubah acak θ. Jika I sukar ditentukan nilainya dengan pengintegralan biasa, maka I dapat didekati nilai dengan *Monte Carlo Integration* yaitu,

$$I = \hat{E}\big(h(\theta)\big) = \frac{1}{n}\sum_{t=1}^{n} h(\theta^t) \tag{2}$$

Analisis Bayesian

Analisis Bayesian menggunakan dua jenis informasi dalam suatu model peluang. Sumber informasi yang pertama berasal dari sampel dan disebut informasi sampel. Informasi sampel di peroleh darisuatu fungsi likelihood karena suatu fungsi likelihood telah memuat informasi tentang parameter yang terkandung di dalam sampel. Informasi prior yang didapat dari opini sang ahli tentang parameter akan bersifat personal atau subyektif.

$$f(x|\theta) = f(x_1|\theta).f(x_2|\theta) \dots .f(x_n|\theta) \tag{3}$$

Sumber informasi kedua berasal dari opini ahli dan disebut informasi prior. Gabungan dua sumber informasi ini akan membentuk informasi posterior yang dinyatakan dalam formula berikut,

$$f(x,\theta) = f(x_1|\theta)f(\theta) \tag{4}$$

$$f(\theta|x) = \frac{f(x,\theta)}{f(x)} \tag{5}$$

Informasi posterior dinyatakan dalam distribusi posterior. Untuk selanjutnya, dari distribusi posterior dapat ditentukan distribusi prediktif $f(y|x)$.

Dalam kerangka analisis Bayesian, parameter diperlakukan sebagai variabel acak. Keacakan parameter bukan berasal dari hasil suatu eksperimen yang berbeda-beda. Keacakan parameter terkait dengan distribusi peluang subyektif yang menggambarkan pengetahuan seseorang tentang parameter tersebut. Hal ini berbeda dengan konsep parameter yang nilainya tetap didalam teori statistik klasik sehingga parameter tidak memiliki distribusi peluang, atau dapat dikatakan parameter memiliki suatu nilai pasti tetapi belum diketahui (klugman dkk, 2004).

Simulasi *Markov Chain Monte Carlo*

Suatu proses stokastik $\{X_0, X_1, X_3, \dots . X_{t, \dots .}\}$ memiliki sifat Markov, jika distribusi peubah acak X_{t+1} bersyarat pada $X_0, X_1, X_3, \dots . X_t$ adalah sama dengan distribusi X_{t+1} bersyarat pada X_t saja. Suatu proses dengan sifat Markov tersebut dinamakan proses Markov. Jika ruang keadaan pada proses markov tersebut terhitung, maka proses Markov tersebut dapat dinyatakan sebagai rantai Markov.

Markov Chain Monte Carlo adalah suatu metode untuk membangkitkan variable acak yang didasarkan pada rantai markov. Metode ini berkaitan dengan estimasi parameter pada inferensi Bayesian. Konsep utama dalam MCMC adalah membuat sampel pendekatan dari distribusi posterior parameter, dengan membangkitkan sebuah rantai Markov yang memiliki distribusi limit mendekati distribusi posterior parameter (Meyers, 2015).

Dengan MCMC akan diperoleh suatu barisan sampel acak yang berkorelasi, yakni nilai ke-j dari barisan U_j disampling dari suatu distribusi peluang yang bergantung pada nilai sebelumnya. Distribusi eksak dari (U_j) umumnya tidak diketahui, namun distribusi pada setiap iterasi dalam barisan nilai sampel tersebut akan konvergen pada distribusi yang sesungguhnya untuk nilai t yang cukup besar. Oleh karena itu, jika ukuran sampel yang diperbaharui cukup besar maka kelompok terakhir dari nilai yang disampling dalam barisan tersebut akan mendekati sebuah sampel yang berasal dari distribusi yang diinginkan.

Gibbs Sampling adalah salah satu algoritma yang terdapat dalam simulasi MCMC. Konsep utama dalam Gibbs sampling adalah menemukan bentuk distribusi bersyarat univariat dimana dalam distribusi tersebut memuat semua variable random dengan satu variabel yang akan ditentukan nilainya. Gibbs sampling memerlukan distribusi bersyarat penuh (full conditional distribution) dari tiap- tiap variabel. Pada Gibbs sampling semua simulasi adalah univariat dan semua sampel hasil simulasi diterima. Gibar sampling bisa diterapkan apabila distribusi probabilitas bersama (joint probability distribution) tidak diketahui secara eksplisit, tetapi distribusi bersyarat penuh (full conditional distribution) dari tiap-tiap variabel diketahui.

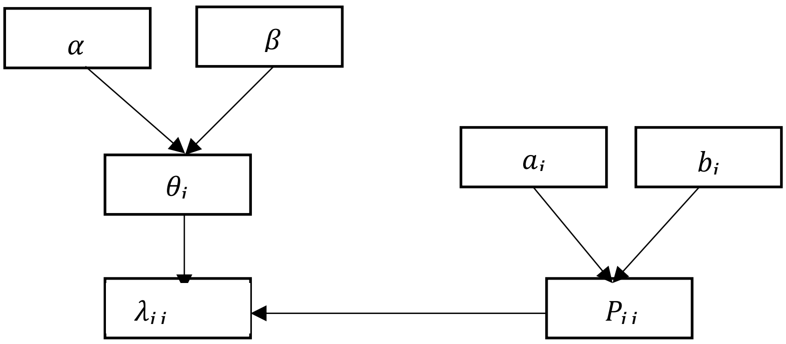
HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini adalah data banyak klaim pada asuransi kompensasi tenaga kerja. Data dibagi dalam tiga grup polis dan lima tahun polis.

Tabel 1. Data Simulasi

	Grup 1		Grup 2		Grup 3	
Tahun	Gaji (dalam puluhan ribu)	Banyak Klaim	Gaji (dalam puluhan ribu)	Banyak Klaim	Gaji (dalam puluhan ribu)	Banyak Klaim
1	269	9	265	7		
2	315	7	255	3	135	7
3	275	6	230	3	125	2
4	320	13	285	7	115	3
5	?	?	275	?	105	?

Beberapa asumsi yang digunakan pada penelitian ini adalah: (1) X_{ij} adalah peubah acak yang menyatakan banyaknya klaim untuk grup polis ke i dan tahun polis ke j. $X_{ij}|\lambda_{ij}$ diasumsikan berdistribusi Poisson (λ_{ij}). (2)



Gambar 1. *Hierarchical Poisson Model*

Konstruksi Model

Setiap sel banyak klaim yang ditampilkan pada Tabel 1, diasumsikan saling bebas sehingga distribusi modelnya adalah perkalian diantara distribusi masing – masing sel banyak klaim. Pada penelitian ini setiap sel diasumsikan mengikuti distribusi Poisson bersayarat $P_{ij}\theta_i$.

$$f\left(\boldsymbol{X}|P_{ij}\theta_i\right) = \prod_{j=1}^{4} f\left(X_{1j}|P_{1j}\theta_1\right) \prod_{j=1}^{4} f\left(X_{2j}|P_{2j}\theta_2\right) \prod_{j=1}^{4} f\left(X_{3j}|P_{3j}\theta_3\right) \tag{6}$$

$$= \prod_{j=1}^{4} \frac{e^{-P_{1j}\theta_1} P_{1j}\theta_1{}^{X_{1j}}}{X_{1j}!} \prod_{j=1}^{4} \frac{e^{-P_{2j}\theta_2} P_{2j}\theta_2{}^{X_{2j}}}{X_{2j}!} \prod_{j=1}^{4} \frac{e^{-P_{3j}\theta_3} P_{3j}\theta_3{}^{X_{3j}}}{X_{3j}!} \tag{7}$$

Berdasarkan gambar 1, terdapat dua tingkat distribusi prior. Distribusi prior pada tingkat pertama adalah distribusi peubah acak θ_i. Sedangkan distribusi prior pada tingkat kedua adalah distribusi peubah acak α dan β. Kedua peubah acak ini diasumsikan saling bebas satu sama lain, sehingga diperoleh

$$f(\theta_i, \alpha, \beta) = \prod_{i=1}^{4} f(\theta_i|\alpha, \beta) f(\alpha) f(\beta) \tag{8}$$

$$f(X, \boldsymbol{\theta}, \alpha, \beta) = \prod_{j=1}^{4} f\left(X_{1j}|P_{1j}\theta_1\right) \prod_{j=1}^{4} f\left(X_{2j}|P_{2j}\theta_2\right) \prod_{j=1}^{4} f\left(X_{3j}|P_{3j}\theta_3\right) \prod_{i=1}^{4} f(\theta_i|\alpha, \beta) f(\alpha) f(\beta) \tag{9}$$

Sedangkan distribusi posterior yang digunakan dalam simulasi ini adalah distribusi posterior bersyarat penuh.

$$f(\theta_1|X, \theta_2, \theta_3, \alpha, \beta) \propto \prod_{j=1}^{4} f\left(X_{1j}|P_{1j}\theta_1\right) \prod_{i=1}^{4} f(\theta_1|\alpha, \beta) \tag{10}$$

$$f(\theta_1|X, \theta_2, \theta_3, \alpha, \beta) \propto \theta^{\alpha + \sum_{j=1}^{4} x_{1j}^{-1}} \exp\left(-\left(\beta + \sum_{j=1}^{4} P_{1j}\right)\theta_1\right) \tag{11}$$

$$\theta_1|X,\theta_2,\theta_3,\alpha,\beta \sim gamma\ (\alpha+\sum_{j=1}^{4} x_{1j}^{-1}, \beta+\sum_{j=1}^{4} P_{1j})$$

$$f(\theta_2|X,\theta_1,\theta_3,\alpha,\beta) \propto \theta^{\alpha+\sum_{j=1}^{4} x_{2j}^{-1}} \exp\left(-\left(\beta+\sum_{j=1}^{4} P_{2j}\right)\theta_2\right) \quad (12)$$

$$\theta_2|X,\theta_1,\theta_3,\alpha,\beta \sim gamma\ (\alpha+\sum_{j=1}^{4} x_{2j}^{-1}, \beta+\sum_{j=1}^{4} P_{2j})$$

$$f(\theta_3|X,\theta_1,\theta_2,\alpha,\beta) \propto \theta^{\alpha+\sum_{j=1}^{4} x_{3j}^{-1}} \exp\left(-\left(\beta+\sum_{j=1}^{4} P_{3j}\right)\theta_3\right) \quad (13)$$

$$\theta_3|X,\theta_1,\theta_2,\alpha,\beta \sim gamma\ (\alpha+\sum_{j=1}^{4} x_{3j}^{-1}, \beta+\sum_{j=1}^{4} P_{3j})$$

$$f(\alpha|X,\theta_1,\theta_2,\theta_3,\beta) \propto \prod_{i=1}^{4} f(\theta_i|\alpha,\beta) f(\alpha) f(\beta) \quad (14)$$

$$\propto \left\{\frac{\beta^\alpha}{\Gamma(\alpha)}\right\}^3 \left\{\prod_{j=2}^{4} \theta_i\right\}^\alpha \alpha^4 \exp(-5\alpha)$$

$$f(\beta|X,\theta_1,\theta_2,\theta_3,\alpha\ ) \propto \beta^{3\alpha+24} \exp\left(-\left(\sum_{i=1}^{3}\theta_i+1\right)\beta\right) \quad (15)$$

$$\beta|X,\theta_1,\theta_2,\theta_3,\alpha \sim gamma\ (3\alpha+24, \sum_{1=1}^{3}\theta_i+1)$$

Untuk menentukan distribusi prediktif, terlebih dahulu dilakukan estimasi terhadap distribusi $f(\theta_i|X)$ menggunakan integral Monte Carlo

$$f(\theta_i|X) = \iint f(\theta_i|X,\alpha,\beta) f(\alpha,\beta|X) d\alpha d\beta$$

$$\approx \frac{1}{n-m} \sum_{t=m+1}^{n} f\left(\theta_i|X,\alpha^{(t)},\beta^{(t)}\ \right) \quad (16)$$

Untuk kemudian diperoleh distribusi prediktof,

$$f(X_{i5}|X) = \int_{\theta_i 1}^{\theta_i 2} f(X_{i5}|P_{i5}\theta_i) f(\theta_i|X)\, d\theta_i$$

$$\approx \frac{1}{n-m} \sum_{t=m+1}^{n} f(X_{i5}|P_{i5}\theta_i^t) \quad (17)$$

Simulasi MCMC

Setelah konstruksi model dilakukan, simulasi untuk memprediksi klaim dapat dijalankan menggunakan algoritma Gibbs Sampling sebagai berikut,

1. Tentukan nilai $\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)}, \alpha^{(0)}, \beta^{(0)}$
2. Untuk iterasi pertama, $t = 1$, bangkitkan nilai dari peubah acak dengan distribusi bersyarat penuh sebagai berikut:
   Bangkitkan $\theta_1^{(1)} \sim f\left(\theta_1|\theta_2^{(0)}, \theta_3^{(0)}, \alpha^{(0)}, \beta^{(0)}\right)$

Bangkitkan $\theta_2^{(1)} \sim f\left(\theta_2 | \theta_1^{(0)}, \theta_3^{(0)}, \alpha^{(0)}, \beta^{(0)}\right)$
Bangkitkan $\theta_3^{(1)} \sim f\left(\theta_3 | \theta_1^{(0)}, \theta_2^{(0)}, \alpha^{(0)}, \beta^{(0)}\right)$
Bangkitkan $\alpha^{(1)} \sim f\left(\alpha | \theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)}, \ \beta^{(0)}\right)$
Bangkitkan $\beta^{(1)} \sim f\left(\beta | \theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)}, \alpha^{(0)}\right)$

3. Dari iterasi pada langkah kedua, diperoleh nilai dari barisan peubah acak$\theta_1^{(1)}, \theta_2^{(1)}, \theta_3^{(1)}, \alpha^{(1)}, \beta^{(1)}$.
4. Langkah kedua diulangi untuk iterasi $t = 1,2,3, \dots, n$ menggunakan nilai yang diperoleh pada langkah ketiga
5. Proses burn in dapat dilakukan jika n sampel yang diperoleh belum konvergen pada suatu distribusi yang ditentukan sebelumnya.

Pada penelitian ini, dilakukan simulasi MCMC dengan menjalankan empat rantai, masing – masing rantai menjalankan 30000 iterasi. Setelah itu dilakukan diagnosa konvergensi sampel dengan mengamati pola *history trace, auto-correlation, MC error* dan BGR Diagnostic. Berdasarkan analisis terhadap empat komponen tersebut, diperoleh beberapa hasil diagnosa sebagai berikut:

1. Plot *history/trace* menunjukkan pola *white noise* dan tidak memiliki trend tertentu
2. *Auto correlation* menunjukkan pola menurun dan mendekati nol
3. MC error menunjukkan angka kurang dari 5% dari standar deviasi
4. BGR *ratio* menunjukkan angka mendekati 1 sejak sampel ke 20001
5. Dilakukan proses burn in pada 20000 sampel pertama, m = 20000
6. Diambil 10000 sampel terakhir dari masing – masing rantai, sehingga total sampe; yang digunakan dalam proses inferensi berjumlah 40000 sampel

Berikut adalah tabel statistik hasil simulasi MCMC pada *Hierarchical Poisson Model* menggunakan *software* OpenBUGS.

Tabel 2. Output software OpenBUGS

Node	Mean	SD	MC Error	2,5%	Median	97.5%	Start	Sample
Alpha	1.027	0.325	0.001549	0.4817	1.024	1.814	20001	40000
Beta	35.49	4.873	0.02715	17.04	25.14	36.18	20001	40000
theta[1]	0.02918	0.004794	2.41E-05	0.02017	0.2912	0.03937	20001	40000
theta[1]	0.01895	0.004794	2.09E-05	0.01117	0.1912	0.02917	20001	40000
theta[1]	0.04	0.004794	5.41E-05	0.02217	0.3912	0.06937	20001	40000
x[1,5]	8.659	3.55	0.01798	3	8	16	20001	40000
x[2,5]	5.554	2.591	0.01257	1	5	11	20001	40000
x[3,5]	4.554	2.696	0.01255	1	4	10	20001	40000

Berdasarkan Tabel 2, diperoleh nilai estimasi dari parameter $\alpha, \beta, \theta_1, \theta_2, \theta_3$ masing – masing bernilai 1.027, 25.49, 0.02918, 0.01895, 0.04. Nilai parameter tersebut dapat digunakan untuk memprediksi banyak klaim pada tahun kelima dengan persamaan,

$$E(X_{i5}|X) = \int E(X_{i5}|P_{i5}\theta_i) f(\theta_i|X) d\theta_i$$
$$\approx \frac{1}{n-m} \sum_{t=m+1}^{n} P_{i5}\, \theta_i{}^{(t)} \tag{18}$$

Rata – rata prediksi banyak klaim pada tahun kelima untuk grup polis 1, 2 dan 3 atau dinyatakan dalam X_{15}, X_{25} dan X_{35} secara berurutan dapat dilihat pada Tabel 2 untuk node x[1,5], x[2,5] dan x[3,5]. Untuk grup polis 1, pada tahun kelima diprediksi terjadi 8.659 klaim atau 9 klaim. Untuk grup polis 2, pada tahun kelima diprediksi terjadi lebih sedikit klaim jika dibandingkan dengan grup 1 yaitu 5.554 klaim. Sedangkan pada grup 3, diprediksi akan terjadi 4.554 klaim.

KESIMPULAN

Pendekatan Bayesian dan simulasi Markov Chain Monte Carlo (MCMC) dapat digunakan untuk memodelkan permasalahan asuransi terkait prediksi banyak klaim di masa yang akan datang. Estimasi banyak klaim dapat dihitung dengan algoritma Gibbs Sampling dengan bantuan software OpenBUGS berdasarkan model yang sudah ditentukan.

DAFTAR RUJUKAN

Best, N.G,. (2000). Bayesian Analysis of Realistically Complex Model. *Journal of The Royal Statistically Society*, pp 323-342.

Cowles, M.K., (2013). *Applied Bayesian Statistics with R and OpenBUGS*. New York: Springer

Fishman, S.G,. (1995). *Monte Carlo Concepts, Alghorithms and Applications*. New York: Springer

Geman, S & Geman, D. (1992). Stochastic Relaxation, Gibbs Distribution and Bayesian Restoration of Image. I*EEE Transaction Pattern Analysis and Machine Intelligence*, (6) pp 721-741.

Klugman, S.A & Panjer H.H (2004). *Loss Models from Data to Decision*. New York: Wiley

Mayers, G. (2015). Stochastic Loss Reserving Using Bayesian MCMC Models. *Casualty Actuarial Society*

Scollnik, D.P.M. (2000). Actuarial Modeling with MCMC and BUGS. *North American Journal*. 5(2), pp 96 – 125.

Scollnik, D.P.M. (1996). An Introduction to Markov Chain Monte Carlo Methods and Their Actuarial Application, Procee*ding of Casualty Actuarial Society.*