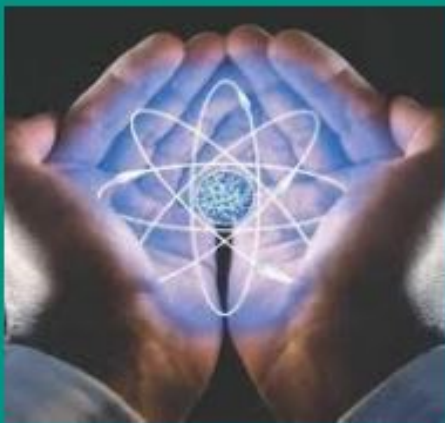


Table Of Content

Journal Cover	2
Author[s] Statement	3
Editorial Team	4
Article information	5
Check this article update (crossmark)	5
Check this article impact	5
Cite this article	5
Title page	6
Article Title	6
Author information	6
Abstract	6
Article content	7

Academia Open



By Universitas Muhammadiyah Sidoarjo

Originality Statement

The author[s] declare that this article is their own work and to the best of their knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the published of any other published materials, except where due acknowledgement is made in the article. Any contribution made to the research by others, with whom author[s] have work, is explicitly acknowledged in the article.

Conflict of Interest Statement

The author[s] declare that this article was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright Statement

Copyright © Author(s). This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>

EDITORIAL TEAM

Editor in Chief

Mochammad Tanzil Multazam, Universitas Muhammadiyah Sidoarjo, Indonesia

Managing Editor

Bobur Sobirov, Samarkand Institute of Economics and Service, Uzbekistan

Editors

Fika Megawati, Universitas Muhammadiyah Sidoarjo, Indonesia

Mahardika Darmawan Kusuma Wardana, Universitas Muhammadiyah Sidoarjo, Indonesia

Wiwit Wahyu Wijayanti, Universitas Muhammadiyah Sidoarjo, Indonesia

Farkhod Abdurakhmonov, Silk Road International Tourism University, Uzbekistan

Dr. Hindarto, Universitas Muhammadiyah Sidoarjo, Indonesia

Evi Rinata, Universitas Muhammadiyah Sidoarjo, Indonesia

M Faisal Amir, Universitas Muhammadiyah Sidoarjo, Indonesia

Dr. Hana Catur Wahyuni, Universitas Muhammadiyah Sidoarjo, Indonesia

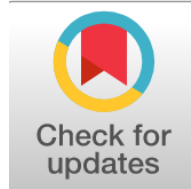
Complete list of editorial team ([link](#))

Complete list of indexing services for this journal ([link](#))

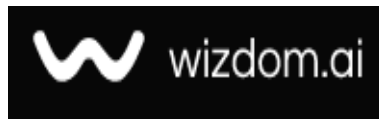
How to submit to this journal ([link](#))

Article information

Check this article update (crossmark)



Check this article impact ^(*)



Save this article to Mendeley



^(*) Time for indexing process is various, depends on indexing database platform

Prediction of Kidney Failure and Cancer Insurance Claims with Bayesian MCMC

Prediksi Klaim Asuransi Gagal Ginjal dan Kanker dengan Bayesian MCMC

Nirmala Ayuningtyas, nirmala.ayuningtyas@students.uui.ac.id, (1)

Program Studi Statistika, Universitas Islam Indonesia , Indonesia

Abdullah Ahmad Dzikrullah, nirmala.ayuningtyas@students.uui.ac.id, (0)

Program Studi Statistika, Universitas Islam Indonesia , Indonesia

⁽¹⁾ Corresponding author

Abstract

General Background: Health insurance plays a crucial role in mitigating the financial risks of catastrophic illnesses. **Specific Background:** In Indonesia, chronic kidney disease (CKD) and cancer contribute significantly to the burden of BPJS Health claims, with rising costs reported in recent years. **Knowledge Gap:** Existing claim estimation models often fail to capture the uncertainty and variability inherent in real-world data. **Aims:** This study aims to develop a Bayesian model with a Markov Chain Monte Carlo (MCMC) approach to accurately estimate insurance claims for CKD and cancer. **Results:** Using 2021–2024 data from RSUP Dr. Soeradji Tirtonegoro Klaten, the model successfully estimated outpatient CKD claims at 1649.29 (SD = 19.82) and outpatient cancer claims at 147.68 (SD = 10.18). All model diagnostics indicate strong convergence and accuracy (R-hat = 1.0, ESS > 5000). **Novelty:** This research applies MCMC-based Bayesian inference with various prior settings (informative to non-informative) and demonstrates robust posterior prediction under different assumptions. **Implications:** The model provides a credible framework for insurance risk management, improving claim prediction and fiscal planning for health providers and insurers, particularly in managing high-cost diseases within the national health system.

Highlight :

- The Bayesian MCMC model produced accurate and stable estimates of kidney failure and cancer claims (R-hat = 1.0, ESS > 5000).
- Sensitivity analysis showed the results remained stable despite different priors, indicating a robust model.
- The best prediction in outpatient CKD (MAPE 5.15%), but less accurate in outpatient cancer (MAPE 50.26%).

Keywords : Bayes, Claims, MCMC, PyMC, Python

Published date: 2025-07-14 00:00:00

Pendahuluan

Asuransi kesehatan merupakan instrumen strategis dalam mitigasi risiko kesehatan, khususnya bagi masyarakat peserta Program Jaminan Kesehatan Nasional (JKN) seperti BPJS Kesehatan. Di antara tantangan utama dalam sistem ini adalah tingginya beban klaim dari penyakit kritis, terutama gagal ginjal kronis (GGK) dan kanker. Berdasarkan Riskesdas 2018, prevalensi GGK di Indonesia mencapai 3,8% dan kanker 1,8%. Data Kementerian Kesehatan tahun 2023 menunjukkan bahwa sepanjang 2022 tercatat lebih dari 1 juta kasus GGK dan 3 juta kasus kanker, dengan total biaya perawatan masing-masing mencapai Rp2,2 triliun dan Rp4,5 triliun.

Penyakit-penyakit ini bukan hanya menjadi beban klinis, tetapi juga menimbulkan tekanan finansial yang serius terhadap program JKN. WHO mencatat kanker sebagai penyebab kematian tertinggi kedua secara global dengan 9,3 juta kematian, sementara penyakit ginjal akibat diabetes menempati peringkat keempat dengan 2 juta kematian. Kondisi serupa tercermin di Indonesia, di mana BPJS Kesehatan melaporkan bahwa penyakit katastropik seperti jantung, stroke, GGK, dan kanker menjadi kontributor utama pembengkakan biaya klaim hingga puluhan triliun rupiah per tahun. Data Dewan Jaminan Sosial Nasional (2023) mengonfirmasi peningkatan signifikan beban jaminan pada penyakit kritis. Dalam kurun 2018–2023, kasus kanker pada peserta JKN meningkat 94%, dengan beban klaim naik 101%; sedangkan GGK mengalami kenaikan beban sebesar 38%. Di RSUP dr. Soeradji Tirtonegoro Klaten, laporan menunjukkan bahwa dalam empat tahun terakhir, klaim asuransi atas GGK dan kanker meningkat secara signifikan. Fakta ini menegaskan pentingnya pengembangan model estimasi yang akurat dalam memprediksi klaim masa depan guna menjaga keberlanjutan fiskal program JKN dan efisiensi layanan rumah sakit.

Kelemahan dalam sistem estimasi klaim saat ini adalah penggunaan metode konvensional yang kurang mampu merepresentasikan ketidakpastian distribusi data klaim secara akurat. Ketidaktepatan estimasi dapat berdampak pada alokasi dana yang tidak optimal, pelayanan yang terganggu, atau peningkatan premi yang memberatkan masyarakat. Oleh karena itu, dibutuhkan pendekatan analitik yang mampu menangkap kompleksitas dan variabilitas klaim asuransi secara lebih realistis. Penelitian ini bertujuan menerapkan pendekatan Bayesian dengan metode *MarkovChainMonte Carlo* (MCMC) dalam memodelkan dan mengestimasi jumlah klaim asuransi atas penyakit GGK dan kanker. Pendekatan ini diharapkan mampu memberikan estimasi distribusi klaim yang lebih fleksibel, realistis, dan informatif bagi pemangku kebijakan.

Pendekatan Bayesian menawarkan keunggulan dalam mengintegrasikan informasi sebelumnya (prior) dengan data terbaru (*likelihood*) untuk menghasilkan distribusi posterior yang merepresentasikan ketidakpastian estimasi. Dalam konteks ini, MCMC digunakan untuk menyimulasikan sampel dari distribusi posterior melalui teknik seperti *Gibbs Sampling*, *Metropolis-Hastings*, dan *HamiltonianMonte Carlo* (HMC) dengan *No-U-TurnSampler* (NUTS). Teknik ini sangat sesuai ketika distribusi analitik tidak dapat dihitung secara eksplisit karena kompleksitas struktur data. Penelitian ini memanfaatkan data klaim riil dari RSUP dr. Soeradji Tirtonegoro Klaten untuk dua jenis penyakit kritis tersebut. Model estimasi dikembangkan menggunakan library Python PyMC dan dievaluasi melalui ArviZ untuk keakuratan hasil dan diagnostik model. Proses estimasi dilakukan dalam sejumlah iterasi untuk membentuk rantai Markov yang konvergen, sehingga mampu memetakan distribusi kemungkinan jumlah klaim secara menyeluruh.

Hasil simulasi dari model MCMC akan menggambarkan sebaran probabilitas dari jumlah klaim secara lebih akurat dibanding metode konvensional. Distribusi posterior yang dihasilkan tidak hanya memberikan nilai estimasi titik, tetapi juga rentang kepercayaan dan struktur ketidakpastian yang diperlukan dalam perencanaan keuangan. Analisis juga memperlihatkan bahwa klaim atas GGK dan kanker memiliki sebaran *overdispersed*, yang tidak bisa ditangani optimal dengan pendekatan frekuentis sederhana. Dengan menggunakan pendekatan ini, BPJS Kesehatan dan rumah sakit dapat melakukan perencanaan dana lebih presisi, penentuan premi yang proporsional, dan pengambilan keputusan berbasis data. Selain itu, model ini dapat disematkan dalam sistem informasi manajemen risiko secara real-time, sehingga mendukung efisiensi operasional dan prediksi keuangan jangka panjang.

Metode Bayesian dengan pendekatan MCMC menawarkan solusi efektif dalam mengatasi ketidakpastian dan kompleksitas data klaim asuransi penyakit kritis. Model ini tidak hanya berguna secara praktis dalam pengelolaan keuangan dan risiko bagi BPJS maupun rumah sakit, tetapi juga berkontribusi terhadap pengembangan aktuarial digital dan sistem prediksi berbasis data. Penerapan tools seperti PyMC dan ArviZ memperkuat validitas model, dan membuka peluang integrasi dalam sistem pembelajaran manajemen risiko serta aktuarial modern. Dengan demikian, penelitian ini memberi nilai tambah strategis dalam konteks keberlanjutan sistem jaminan kesehatan nasional.

Metode

Fungsi Likelihood

Fungsi *likelihood* didefinisikan sebagai fungsi densitas bersama dari variabel acak $X_1, X_2, \dots, X_n$ dan dinyatakan dalam bentuk $f(x_1, x_2, \dots, x_n | \theta)$. Jika nilai $x_1, x_2, \dots, x_n$ sudah tetap, maka fungsi *likelihood* menjadi fungsi dari parameter θ , yang biasanya dinotasikan sebagai $L(\theta)$. Ketika $x_1, x_2, \dots, x_n$ merupakan sampel acak dari distribusi

$f(x|\theta)$, maka fungsi *likelihood* dapat ditulis sebagai :

$$L(\theta|x) = f(x_1|\theta)f(x_2|\theta) \dots f(x_n|\theta) = \prod_{i=1}^n f(x_i|\theta) \quad (1)$$

Figure 1.

Penting untuk dicatat bahwa fungsi *likelihood* pada persamaan 1 ini hanyalah fungsi dengan parameter , yang nilainya belum diketahui. Estimator maksimum *likelihood* adalah nilai yang memaksimalkan fungsi *likelihood* .

Ragam dan Standar Deviasi

Ragam adalah suatu ukuran statistik yang digunakan untuk mengukur variasi antara data pengamatan serta menggambarkan distribusi dari data tersebut. Sedangkan standar deviasi mencerminkan ukuran standar untuk kelompok data yang dianalisis. Nilai standar deviasi mengikuti satuan yang digunakan dalam pengukuran data. Jika setiap data memiliki acuan pada harga rata-ratanya $\bar{x}$, maka akan diperoleh standar deviasi sebesar $d = x_i - \bar{x}$.

Ragam populasi, untuk populasi berhingga dengan elemen $x_1, x_2, x_3, \dots, x_n$ dirumuskan sebagai:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{N} \quad (2)$$

Dimana μ adalah nilai rata-rata untuk populasi. Sedangkan ragam sampel berhingga $x_1, x_2, x_3, \dots, x_n$ dirumuskan sebagai:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (3)$$

Nilai $\bar{x}$ adalah nilai rata-rata untuk sampel. Standar deviasi merupakan akar dari ragam atau variansi. Standar deviasi suatu distribusi sampel statistik dikenal juga sebagai galat baku dari statistik yang dihitung [18].

Figure 2.

Metode Bayes

Seorang matematikawan Inggris, Thomas Bayes, mengembangkan teori untuk menghitung probabilitas suatu penyebab berdasarkan hasil pengamatan, yang dikenal sebagai Teorema Bayes. Metode ini dikembangkan oleh Thomas Bayes pada abad ke-18 dan telah menjadi dasar dari banyak aplikasi dalam ilmu data, *machine learning*, deteksi spam, prediksi cuaca, diagnosis medis, dan banyak lagi . Teori ini memungkinkan penggabungan informasi tambahan untuk memperbaiki probabilitas awal (prior) menjadi probabilitas setelah data baru diperoleh (posterior) .

Pada bidang aktuaria, metode Bayesian dimanfaatkan untuk memperbarui prediksi estimasi terhadap kerugian dimasa mendatang. Misalnya isalnya, jika menyatakan variabel acak yang merepresentasikan kerugian (seperti jumlah klaim, besar klaim, atau total kerugian), maka distribusi dari akan bergantung pada suatu parameter . Karena parameter ini dapat berubah, maka dianggap sebagai variabel acak dengan distribusinya sendiri (Tse, 2014).

Dalam pendekatan Bayesian, parameter diperlakukan sebagai variabel acak, dan distribusi posterior dihitung dengan cara menggabungkan informasi awal (prior) dan informasi yang berasal dari data (*likelihood*). Langkah pertama dalam pendekatan ini adalah menentukan distribusi prior. Menurut Klugman (2019), prior mencerminkan keyakinan awal terhadap parameter berdasarkan informasi atau pengetahuan sebelumnya. Prior ini bisa berupa prior informatif, yang mencerminkan pengetahuan awal yang kuat, atau prior non-informatif, yang digunakan ketika informasi sebelumnya sangat terbatas.

Distribusi Prior san Posterior

Dalam pendekatan metode Bayes, parameter suatu distribusi dianggap sebagai variabel acak yang memiliki distribusi tertentu. Misalnya , maka parameter tersebut juga dapat dianggap sebagai variabel acak yang memiliki distribusi probabilitas tertentu, yang disebut sebagai distribusi prior. Distribusi ini biasanya dilambangkan dengan , sehingga dapat ditulis sebagai . Pemilihan fungsi densitas distribusi prior seringkali menjadi tantangan. Dalam beberapa kasus, parameter prior ini diasumsikan sebagai variabel acak, baik yang berbentuk diskrit maupun kontinu . Distribusi prior ini mencerminkan tingkat keyakinan peneliti terhadap parameter yang akan diestimasi berdasarkan pengetahuan dan pengamatan sebelumnya.

Metode Bayes menggabungkan distribusi prior dengan distribusi bersyarat dari data sampel , untuk menghitung sebaran distribusi posterior , sehingga dapat dikatakan bahwa distribusi posterior merupakan distribusi bersyarat dari parameter setelah data sampel diketahui. Distribusi posterior ini mendefinisikan keyakinan peneliti terhadap parameter setelah data sampel diamati . Secara umum, statistika Bayes memiliki kerangka kerja sebagai berikut:

1. Keyakinan awal yang diambil dari bermacam-macam kemungkinan mengikuti asumsi sebuah sebaran prior untuk parameter yang diinginkan.
2. Dengan menggunakan distribusi prior dan fungsi *likelihood* dari data yang diamati, distribusi posterior dapat dihitung. Rumus umum untuk distribusi posterior adalah:

$$f(\theta|x) = \begin{cases} \frac{f(x|\theta)f(\theta)}{\int_{-\infty}^{\infty} f(x|\theta)f(\theta)d(\theta)}, & \text{untuk variabel kontinu} \\ \frac{f(x|\theta)f(\theta)}{\sum_{s=-\infty}^{\infty} f(x|\theta)f(\theta)}, & \text{untuk variabel diskrit} \end{cases} \quad (4)$$

Di mana $f(x|\theta)$ adalah fungsi *likelihood* dan $f(\theta)$ adalah distribusi prior. Penyebut dari persamaan 4 tersebut, yaitu $\int_{-\infty}^{\infty} f(x|\theta)f(\theta)d(\theta)$ dan $\sum_{s=-\infty}^{\infty} f(x|\theta)f(\theta)$ adalah suatu konstanta yang menyebabkan $f(x|\theta)$ menjadi fungsi kepadatan peluang yang layak [21].

Figure 3.

Menurut , distribusi prior dapat diklasifikasikan menjadi dua kategori, yaitu berdasarkan bentuk distribusi dan cara penentuan parameternya. Berdasarkan bentuk distribusinya, distribusi prior dibedakan menjadi dua jenis yaitu:

1. Distribusi prior konjugat yaitu, distribusi prior yang memiliki pola sama dengan distribusi data. Misalnya distribusi beta sebagai prior konjugat untuk parameter dalam data yang mengikuti sebaran distribusi binomial, atau distribusi normal sebagai prior konjugat untuk parameter dalam data yang menyebar normal.
2. Distribusi prior non-konjugat yaitu, distribusi prior yang polanya tidak sama atau tidak sesuai dengan distribusi data.

Metode *Markov Chain Monte Carlo* sering digunakan untuk memperkirakan nilai parameter dari distribusi posterior (). Teknik ini sangat efektif dalam menyelesaikan estimasi parameter yang kompleks dan sulit dikerjakan menggunakan metode lain.

Uji Goodness of Fit

Pada penelitian ini, sebelum dilakukan pemodelan Bayesian, dilakukan pengujian awal terhadap karakteristik distribusi data observasi klaim asuransi. Salah satu metode yang digunakan adalah uji normalitas *Kolmogorov-Smirnov*, untuk mengidentifikasi apakah data awal memenuhi asumsi distribusi tertentu. Uji *Goodness of Fit* adalah metode statistik yang digunakan untuk menguji apakah data observasi berasal dari suatu populasi dengan distribusi tertentu, misalnya distribusi normal, eksponensial, atau lainnya, yang dihipotesiskan. Salah satu metode yang sering digunakan untuk melakukan uji ini adalah uji satu sampel *Kolmogorov-Smirnov* (K-S) . Dalam uji *Kolmogorov-Smirnov*, tingkat kesalahan yang dapat diterima biasanya berada dalam rentang . Dengan tingkat kepercayaan (SK) sebesar , maka tingkat kepercayaan dapat bernilai antara hingga , atau dalam interval .

Distribusi Beta

Distribusi Beta adalah distribusi probabilitas kontinu yang didefinisikan pada interval [0, 1]. Distribusi ini sering digunakan untuk menggambarkan distribusi dari variabel acak yang dinyatakan dalam bentuk proporsi atau probabilitas, terutama dalam konteks Bayesian atau analisis proporsi. Distribusi Beta bergantung pada dua parameter bentuk , yang menentukan bentuk kurva distribusi. nilai yang berbeda dari menghasilkan bentuk distribusi yang dapat simetris atau condong ke kiri atau ke kanan. Distribusi Beta juga fleksibel dalam memodelkan berbagai pola kepercayaan, karena ketika lebih besar dari 1, distribusi berbentuk lonceng (mirip normal), sementara untuk nilai yang lebih kecil dari 1, distribusi menjadi lebih curam atau berfokus di sekitar 0 dan 1. Karakteristik ini menjadikannya umum dalam analisis Bayesian sebagai distribusi prior untuk probabilitas dan proporsi.

Jika x adalah variabel acak yang mengikuti distribusi Beta dengan parameter , maka fungsi rata- rata dan variansinya adalah berikut :

$$E(x) = \frac{\alpha}{\alpha + \beta} \quad (5)$$

$$Var(x) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \quad (6)$$

Figure 4.

Distribusi Beta sering digunakan dalam teori Bayesian untuk mendefinisikan prior pada proporsi atau probabilitas, serta dalam berbagai aplikasi lainnya seperti estimasi proporsi dan analisis statistik yang melibatkan proporsi. Untuk menduga parameter pada distribusi Beta menggunakan metode momen, kita dapat menggunakan momen ke-1 (rata-rata) dan momen ke-2 (varians) dari distribusi ini. Sedangkan untuk distribusi prior

dari P beta yaitu $f(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1}$ dimana $0 \leq p \leq 1$, sehingga diperoleh persamaan untuk distribusi Posterior dengan asumsi P berdistribusi Beta (α, β) sebagai berikut [25]:

$$f(p|x, n) \sim \text{Beta}(x + \alpha, n - x + \beta) \quad (7)$$

Figure 5.

Apabila data jumlah klaim diasumsikan mengikuti distribusi prior Beta dengan parameter awal Beta (α, β) , maka setelah diperoleh data aktual berupa jumlah klaim sebanyak x dari total n observasi, distribusi tersebut akan diperbarui menjadi distribusi posterior yang juga berbentuk Beta, namun dengan parameter yang telah disesuaikan. Dalam hal ini, parameter pertama dari distribusi posterior menjadi $x + \alpha$, yang mencerminkan penambahan jumlah klaim aktual ke dalam estimasi awal terhadap keberhasilan (klaim terjadi), sedangkan parameter kedua menjadi $n - x + \beta$, yang menggabungkan jumlah klaim yang tidak terjadi ke dalam estimasi awal terhadap kegagalan. Proses pembaruan ini mencerminkan prinsip dalam inferensi Bayesian, di mana informasi awal (prior) diolah bersama data observasi (*likelihood*) untuk menghasilkan distribusi posterior yang lebih akurat dalam merepresentasikan keyakinan terhadap parameter yang diestimasi, dalam hal ini probabilitas terjadinya klaim.

Distribusi Normal

Distribusi normal merupakan salah satu distribusi probabilitas yang paling penting dalam statistik. Distribusi ini berbentuk simetris dan menyerupai lonceng (*bell-shaped curve*), dengan rata-rata, median, dan modus memiliki nilai yang sama. Distribusi normal digunakan untuk memodelkan berbagai fenomena alami dan sosial, di mana sebagian besar data berkumpul di sekitar nilai tengah dan kemungkinan kejadian ekstrim semakin kecil. menjelaskan bahwa secara matematis, fungsi kepadatan probabilitas dari distribusi normal dengan rata-rata μ dan standar deviasi dinyatakan sebagai berikut:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (8)$$

Figure 6.

Pada konteks analisis statistik, distribusi normal menjadi dasar dalam berbagai metode inferensial seperti uji hipotesis, estimasi parameter, dan regresi. Selain itu, banyak metode simulasi seperti *Markov Chain Monte Carlo* (MCMC) menggunakan distribusi normal sebagai prior, proposal distribution, atau untuk asumsi error model. Pada penelitian ini, pengujian awal terhadap data observasi klaim dilakukan dengan uji normalitas menggunakan metode *Kolmogorov-Smirnov*, untuk mengetahui apakah data mengikuti distribusi normal sebelum dilakukan pemodelan.

Distribusi Half Normal

Distribusi Half Normal adalah distribusi probabilitas yang diperoleh dari distribusi normal dengan hanya mempertimbangkan nilai-nilai non-negatif. Secara matematis, distribusi Half Normal merupakan distribusi normal standar yang dipotong pada nol dan semua nilai negatif direfleksikan menjadi positif. Dengan kata lain, distribusi ini menggambarkan situasi di mana hanya besarnya deviasi dari suatu nilai tengah yang menjadi fokus, tanpa memperhatikan arah deviasi tersebut. Fungsi kepadatan probabilitas dari distribusi Half Normal dengan parameter skala σ dinyatakan sebagai:

$$F(x) = \frac{\sqrt{2}}{\sigma\sqrt{\pi}} e^{-\frac{x^2}{2\sigma^2}}, x \geq 0 \quad (9)$$

Figure 7.

Distribusi ini memiliki rata-rata sebesar $\sigma\sqrt{2/n}$, dengan varians sebesar $\sigma^2(1-2/n)$. Pada konteks analisis Bayesian, distribusi Half Normal sering digunakan sebagai prior untuk parameter standar deviasi (σ) karena sifatnya yang hanya mendukung nilai positif, sehingga sesuai untuk mengasumsikan parameter yang harus bernilai non-negatif. Dalam penelitian ini, distribusi Half Normal digunakan untuk membentuk prior pada parameter deviasi standar klaim asuransi kesehatan [25].

Metode Markov Chain Monte Carlo

Metode *Markov Chains Monte Carlo* (MCMC) adalah pendekatan simulasi yang digunakan untuk menghasilkan sampel dari distribusi variabel acak dengan teknik sampling yang memanfaatkan sifat rantai Markov. MCMC sangat berguna untuk menduga parameter suatu peubah acak dalam sebaran posterior yang rumit jika diselesaikan secara analitik. Metode *Markov Chain Monte Carlo* (MCMC) telah banyak diterapkan di berbagai bidang untuk menyelesaikan beragam masalah, terutama yang berkaitan dengan inferensi Bayesian. Metode ini digunakan untuk menentukan distribusi posterior maupun distribusi prior dalam berbagai studi kasus serta, menghasilkan sampel berulang dari distribusi posterior, yang kemudian digunakan untuk estimasi parameter dan prediksi. Keunggulan MCMC terletak pada fleksibilitasnya dalam menangani model dengan distribusi posterior yang rumit (Speagle, 2019). MCMC dapat diterapkan pada kasus dengan variabel tunggal (univariat) maupun banyak variabel (multivariat). Dua algoritma utama yang sering digunakan dalam MCMC adalah *Metropolis-Hastings* dan *Gibbs Sampling*. Dalam praktik modern, algoritma No-U-Turn Sampler (NUTS) menggantikan Metropolis-Hastings untuk model dengan parameter kontinu karena efisiensinya dalam menjelajahi ruang parameter.

Library PYMC (NUTS) dan Arviz

Penelitian ini menggunakan algoritma *No-U-Turn Sampler* (NUTS) melalui library PyMC di Python untuk inferensi model probabilistik. NUTS adalah metode sampling MCMC yang lebih efisien dibanding *Gibbs Sampling* dan *Metropolis-Hastings*, terutama untuk data dengan korelasi tinggi antar parameter. Berbeda dengan algoritma tradisional yang menggunakan random walk dan memerlukan pengaturan manual langkah, NUTS secara otomatis mengoptimalkan panjang langkah sampling, sehingga mempercepat konvergensi dan mengurangi risiko terjebak di local maxima. Untuk diagnosis dan visualisasi hasil sampling, digunakan library Arviz yang membantu mengevaluasi konvergensi dan distribusi posterior. Secara keseluruhan, pemilihan NUTS meningkatkan efisiensi dan akurasi inferensi pada model dengan data berkorelasi tinggi, sehingga sangat direkomendasikan dibanding metode sampling konvensional.

Tuning Parameter

Penelitian ini menggunakan tuning parameter pada algoritma NUTS di PyMC untuk mengoptimalkan proses sampling. Pengaturan meliputi target acceptance rate 0,9, 1000 iterasi tuning, step size otomatis, serta 2000 draws dengan 4 rantai paralel. Tuning ini memastikan hasil sampling stabil, akurat, dan memenuhi kriteria konvergensi seperti $\hat{r}$ mendekati 1 dan *Effective Sample Size* memadai.

Mean Absolute Percentage Error (MAPE) dan Root Mean Square Error (RMSE)

Pada penelitian ini, evaluasi performa model prediksi jumlah klaim dilakukan menggunakan dua metrik utama, yaitu *Mean Absolute Percentage Error* (MAPE) dan *Root Mean Square Error* (RMSE). MAPE digunakan untuk mengukur tingkat kesalahan prediksi dalam bentuk persentase terhadap nilai aktual, sehingga memberikan gambaran keakuratan model secara relatif. Sementara itu, RMSE mengukur deviasi rata-rata antara nilai prediksi dan aktual dalam satuan yang sama dengan data, sehingga mencerminkan kesalahan secara absolut. Penggunaan kedua metrik ini secara bersamaan memberikan penilaian yang lebih komprehensif: MAPE menunjukkan efisiensi prediktif model dalam konteks proporsional, sedangkan RMSE menilai kestabilan prediksi terhadap penyimpangan nilai besar. Dengan pendekatan ini, evaluasi model menjadi lebih kuat dalam merepresentasikan akurasi estimasi jumlah klaim asuransi berdasarkan metode Bayesian MCMC.

Diagram Alur Penelitian

Adapun diagram alur penelitian disajikan sebagai berikut untuk memberikan gambaran sistematis mengenai tahapan yang dilakukan dalam proses penelitian. Diagram ini memvisualisasikan langkah-langkah utama yang dimulai dari identifikasi masalah, pengumpulan data klaim asuransi penyakit gagal ginjal kronis dan kanker, hingga proses analisis menggunakan pendekatan Bayesian dengan metode *Markov Chain Monte Carlo* (MCMC). Setiap tahap dirancang secara runtut untuk memastikan bahwa proses estimasi dilakukan secara valid, akurat, dan dapat dipertanggungjawabkan baik secara statistik maupun secara praktis dalam konteks pengelolaan klaim asuransi kesehatan.



Figure 8. Diagram Alur Penelitian

Hasil dan Pembahasan

Tahun	Jenis Penyakit	Jenis Layanan	Jumlah Klaim
2021	Gagal Ginjal Kronis	Rawat Inap (RANAP)	1451
		Rawat Jalan (RAJAL)	20982
	Kanker (Kemoterapi)	Rawat Inap (RANAP)	273
		Rawat Jalan (RAJAL)	861
2022	Gagal Ginjal Kronis	Rawat Inap (RANAP)	1702
		Rawat Jalan (RAJAL)	19266
	Kanker (Kemoterapi)	Rawat Inap (RANAP)	393
		Rawat Jalan (RAJAL)	1580
2023	Gagal Ginjal Kronis	Rawat Inap (RANAP)	2218
		Rawat Jalan (RAJAL)	19119
	Kanker (Kemoterapi)	Rawat Inap (RANAP)	498
		Rawat Jalan (RAJAL)	2753
2024	Gagal Ginjal Kronis	Rawat Inap (RANAP)	2459
		Rawat Jalan (RAJAL)	16512
	Kanker (Kemoterapi)	Rawat Inap (RANAP)	371
		Rawat Jalan (RAJAL)	1611
Total	Gagal Ginjal Kronis	Rawat Inap (RANAP)	7830
		Rawat Jalan (RAJAL)	75879
	Kanker (Kemoterapi)	Rawat Inap (RANAP)	1535
		Rawat Jalan (RAJAL)	6805

Table 1. Total Data Klaim Pertahun

Penyakit	Mean	Std Dev	Min	Max	Median	Skewness	Kurtosis
GGK_Rajal	1649,543	128,297	1456	2021	1616,5	1,499111	1,908004
GGK_Ranap	170,2174	54,1307	65	287	166,5	0,238547	-0,580572

Kanker_Rajal	147,9348	68,0279	45	323	136,5	0,532086	-0,349712
Kanker_Ranap	33,36957	9,9294	15	53	33,5	0,175103	-0,882003

Table 2. Hasil Uji Statistika Deskriptif Data Klaim BPJS

Berdasarkan data pada Tabel 1, penelitian menunjukkan total data klaim asuransi untuk dua penyakit, Gagal Ginjal Kronis (GGK) dan Kanker (Kemoterapi), pada pasien di RSUP dr. Soeradji Tirtonegoro Klaten antara Januari 2021 hingga Oktober 2024. Klaim untuk GGK menunjukkan peningkatan konsisten setiap tahunnya, baik untuk Rawat Inap (RANAP) yang naik dari 1.451 klaim pada 2021 menjadi 2.459 klaim pada 2024, maupun untuk Rawat Jalan (RAJAL) meskipun ada penurunan pada 2024, yaitu dari 20.982 klaim di 2021 menjadi 16.512 klaim.

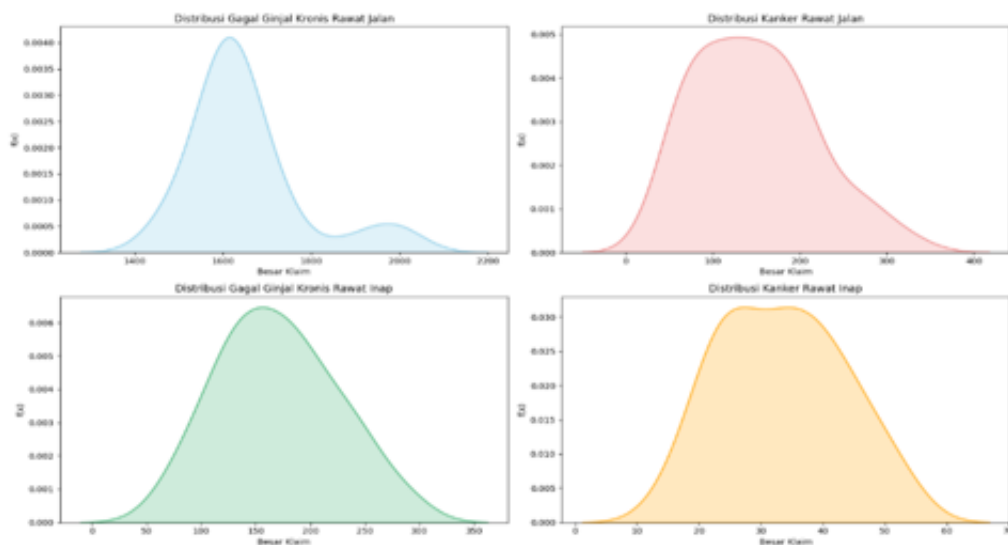
Untuk Kanker, klaim Rawat Inap lebih rendah dibandingkan GGK, dengan fluktuasi tahunan antara 273 klaim pada 2021 hingga 371 klaim pada 2024. Klaim Rawat Jalan untuk Kanker meningkat dari 861 klaim pada 2021 menjadi 1.611 klaim pada 2024. Total klaim untuk GGK mencapai 83.709, terdiri dari 7.830 klaim untuk RANAP dan 75.879 klaim untuk RAJAL. Sementara untuk Kanker, total klaim berjumlah 8.340, dengan 1.535 klaim untuk RANAP dan 6.805 klaim untuk RAJAL.

Sementara itu pada Tabel 2 menunjukkan nilai statistika deskriptif, rata-rata tertinggi dimiliki oleh klaim GGK Rawat Jalan dengan nilai 1649 dan terendahnya dimiliki oleh Kanker Rawat Inap dengan nilai sebesar 33. Sedangkan, Kanker Rawat jalan dan GGK Rawat Inap memiliki rata-rata berkisar antara 140-170 per bulan.

Secara keseluruhan, klaim untuk GGK lebih tinggi dibandingkan Kanker, terutama untuk Rawat Jalan. Tren ini menunjukkan bahwa GGK lebih sering membutuhkan perawatan intensif dan berkelanjutan, sedangkan Kanker menunjukkan peningkatan klaim Rawat Jalan, yang mencerminkan kebutuhan akan perawatan kanker yang lebih sering dilakukan secara rawat jalan.

Hasil Uji Normalitas

Uji normalitas bertujuan untuk mengetahui apakah data klaim asuransi penyakit gagal ginjal kronis (GGK) dan kanker yang dianalisis mengikuti distribusi normal. Pada penelitian ini, uji normalitas dilakukan dengan metode *Kolmogorov-Smirnov*, dan hasilnya disajikan pada Tabel 3.

**Figure 9.** Density Plot Data Klaim Penyakit GGK dan Kanker

Berdasarkan hasil analisis pada Gambar 2, menggunakan *density plot* terlihat bahwa data klaim untuk Gagal Ginjal Kronis dan Kanker, baik untuk rawat jalan maupun rawat inap, cenderung mengikuti distribusi normal. Hal ini ditunjukkan oleh bentuk kurva yang menyerupai lonceng, meskipun pada gagal ginjal kronis dengan jenis pelayanan rawat jalan terdapat sedikit kemiringan (*skewness*) ke kanan. Sebagian besar nilai besar klaim berkumpul di sekitar 1600-1700, dengan beberapa *outlier* di atas 2000. Oleh karena itu, bentuk distribusinya menunjukkan bahwa besar klaim untuk rawat jalan ini relatif konsisten, namun ada variasi ke arah nilai lebih tinggi yang menyebabkan ekor distribusi memanjang.

Jenis Penyakit	Statistic	df	p-value	Keputusan
----------------	-----------	----	---------	-----------

GGK_Rajal	0.2129	46	0.0262	Tolak H ₀
GGK_Ranap	0.0620	46	0.9897	Gagal Tolak H ₀
Kanker_Rajal	0.0841	46	0.8740	Gagal Tolak H ₀
Kanker_Ranap	0.1047	46	0.6557	Gagal Tolak H ₀

Table 3. Hasil Uji Normalitas Kolmogorov-Sminorv Test

Berdasarkan hasil uji normalitas pada Tabel 3, untuk data klaim GGK rawat jalan (GGK_Rajal), diperoleh nilai statistik sebesar 0,2129 dengan p -value sebesar 0,0262. Karena p -value lebih kecil dari taraf signifikansi 0,05, maka keputusan yang diambil adalah menolak hipotesis nol (H_0), yang berarti data GGK_Rajal tidak berdistribusi normal. Sebaliknya, untuk data klaim GGK rawat inap (GGK_Ranap) diperoleh statistik sebesar 0,0620 dengan p -value sebesar 0,9897. Begitu pula untuk data kanker rawat jalan (Kanker_Rajal) dengan p -value sebesar 0,8740, dan kanker rawat inap (Kanker_Ranap) dengan p -value sebesar 0,6557. Ketiga kelompok data ini memiliki p -value lebih besar dari 0,05, sehingga keputusan yang diambil adalah gagal menolak hipotesis nol. Dengan demikian, data klaim GGK_Ranap, Kanker_Rajal, dan Kanker_Ranap dapat disimpulkan berdistribusi normal.

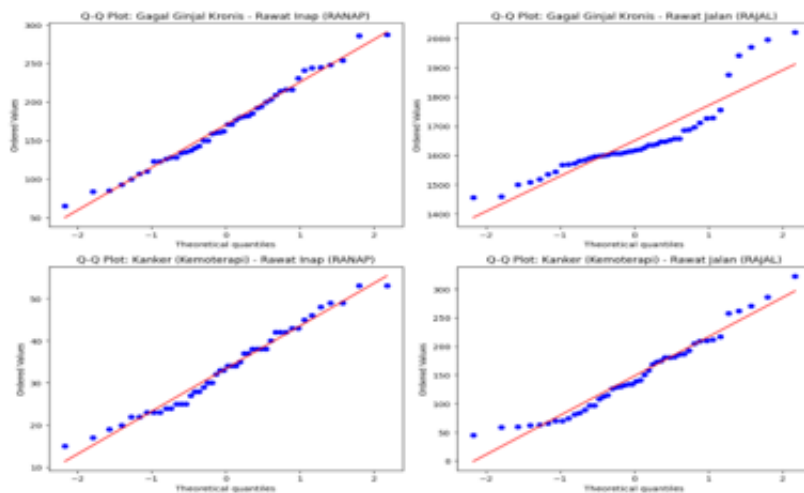


Figure 10. Grafik Korelasi pada Data Klaim Asuransi

Gambaran visual terhadap pola distribusi data disajikan pada Gambar 2. Grafik distribusi peluang tersebut menunjukkan bentuk penyebaran data untuk masing-masing jenis klaim. Selain itu, untuk memperkuat pemahaman mengenai normalitas data, analisis grafik korelasi juga dilakukan, sebagaimana ditampilkan dalam Gambar 3. Grafik korelasi tersebut memberikan indikasi hubungan antara nilai observasi aktual dengan nilai yang diharapkan jika data mengikuti distribusi normal. Pola yang mendekati garis lurus menguatkan hasil uji statistik yang telah diperoleh. Hasil $Q-Q$ plot pada Gambar 3, terlihat bahwa data klaim untuk Gagal Ginjal Kronis dan Kanker, baik untuk rawat jalan maupun rawat inap, cenderung mengikuti distribusi normal. Hal ini ditunjukkan oleh penyebaran titik pada $Q-Q$ plot yang mendekati garis lurus. Grafik $Q-Q$ plot untuk rawat jalan pada gagal ginjal kronis juga menunjukkan kecenderungan mengikuti distribusi normal, namun ada deviasi lebih jelas di bagian *quantile* atas (data besar). Namun, secara keseluruhan data sudah cukup memenuhi asumsi normalitas.

Dengan mempertimbangkan hasil uji normalitas dan visualisasi grafik, dapat disimpulkan bahwa sebagian besar data klaim memenuhi asumsi normalitas. Dengan demikian, karena data klaim (baik gagal ginjal maupun kanker) menunjukkan pola distribusi yang cukup normal namun dengan korelasi tinggi antar parameter, maka algoritma sampling seperti NUTS (*No-U-Turn Sampler*) akan digunakan dalam penelitian ini. NUTS tidak hanya mempercepat konvergensi, tetapi juga menghasilkan estimasi parameter yang lebih akurat dan stabil, bahkan dalam model yang memiliki dependensi kuat antar variabel. Hal ini menjadi perhatian dalam pemilihan metode analisis lanjutan yang sesuai dengan karakteristik distribusi data.

Hasil Implementasi Metode Bayes dengan Model MCMC

Data Awal Asumsi Prior

Pada penerapan metode Bayesian, keakuratan estimasi sangat dipengaruhi oleh pemilihan dan pembaruan parameter distribusi prior menjadi distribusi posterior. Salah satu aspek penting dalam hal ini adalah kejelasan mengenai asal-usul nilai parameter distribusi posterior yang digunakan dalam analisis, khususnya pada model distribusi Beta yang sering digunakan untuk menggambarkan proporsi kejadian. Pada penelitian ini, distribusi Beta digunakan untuk mengestimasi proporsi klaim asuransi terhadap total populasi klaim yang tersedia. Oleh karena

itu, penting untuk menjelaskan secara eksplisit bagaimana nilai parameter α dan β pada distribusi posterior ditentukan, agar tidak menimbulkan keraguan terhadap ketepatan teknis dan metodologis yang diterapkan dalam proses estimasi.

Penentuan parameter α dan β pada distribusi posterior untuk estimasi proporsi klaim penyakit kanker dan gagal ginjal kronis didasarkan secara langsung pada data aktual jumlah klaim yang tercatat di RSUP dr. Soeradji Tirtonegoro Klaten selama periode Januari 2021 hingga Oktober 2024. Berdasarkan rekapitulasi data, jumlah klaim untuk penyakit pada Tabel 1, kanker tercatat sebanyak 8.340 kasus, yang terdiri atas 1.535 klaim rawat inap dan 6.805 klaim rawat jalan. Sementara itu, jumlah klaim untuk penyakit gagal ginjal kronis mencapai 83.709 kasus, dengan rincian 7.830 klaim rawat inap dan 75.879 klaim rawat jalan. Dengan demikian, total seluruh klaim dari kedua penyakit ini selama periode pengamatan berjumlah 92.049 kasus.

Pada penelitian ini, distribusi prior Beta dengan parameter awal $\alpha = 2$ dan $\beta = 10$, yang mencerminkan keyakinan awal dengan bobot informasi yang lemah (*less informative*). Setelah data aktual diamati, distribusi posterior dihitung dengan memperbarui parameter prior berdasarkan informasi dari data menggunakan formula konjugat distribusi Beta, pada persamaan 7, di mana x adalah jumlah klaim aktual dan n adalah total klaim. Untuk kasus kanker, diperoleh distribusi posterior Beta (8342, 83719), dengan $\alpha = 2 + 8340$ dan $\beta = 10 + (92049 - 8340)$. Adapun untuk kasus gagal ginjal kronis, distribusi posterior menjadi Beta (83711, 8342), diperoleh dari $\alpha = 2 + 83709$ dan $\beta = 10 + (92049 - 83709)$. Seperti yang dijelaskan pada Tabel 4 dan 5 berikut:

No.	Kategori	Deskripsi	Nilai
1	Statistik Data		
	Total Pasien	Jumlah keseluruhan pasien yang dianalisis.	92,049
	Total Pasien Kanker	Jumlah pasien kanker dalam data.	8,340
	Proporsi Sampel	Persentase pasien kanker dari total sampel.	9.06% (0.0906)
2	Distribusi Prior		
	Prior untuk p	Distribusi Beta yang digunakan untuk prior.	Beta ($\alpha=2$, $\beta=10$)
	Mean Prior	Nilai rata-rata dari distribusi prior.	0.1667
	Standar Deviasi Prior	Standar deviasi dari distribusi prior.	0.1034
3	Distribusi Posterior		
	Posterior untuk p	Distribusi Beta yang dihasilkan setelah data dianalisis.	Beta ($\alpha=8,342$, $\beta=83,719$)
	Mean Posterior	Rata-rata estimasi proporsi pasien kanker setelah analisis Bayesian.	9.06% (0.0906)
	Standar Deviasi Posterior	Standar deviasi hasil posterior yang menunjukkan ketepatan estimasi.	0
	Interval Kredibel 95%	Rentang nilai yang diperkirakan dengan tingkat kepercayaan 95%.	(0.0888, 0.0925)
4	Interpretasi	Estimasi proporsi pasien kanker adalah 9.06%, dengan interval kredibel 95% antara 8.88% dan 9.25%.	

Table 4. Data Awal Asumsi Prior Pada Kanker

Berdasarkan analisis Bayesian terhadap 92.049 pasien, ditemukan 8.340 penderita kanker (9,06%). Estimasi proporsi populasi dilakukan dengan distribusi prior Beta (2,10), mencerminkan keyakinan awal dengan rata-rata 0,1667 dan standar deviasi 0,1034. Setelah mempertimbangkan data, diperoleh distribusi posterior Beta (8342, 83719) dengan rata-rata 0,0906 dan standar deviasi sangat kecil, menandakan estimasi yang akurat. Interval kredibel 95% berada pada (0,0888; 0,0925), yang menunjukkan keyakinan tinggi bahwa proporsi sebenarnya ada dalam rentang tersebut. Pendekatan Bayesian menghasilkan estimasi proporsi terbaik sebesar 9,06%.

No.	Kategori	Deskripsi	Nilai
1	Statistik Data		

	Total Pasien	Jumlah keseluruhan pasien yang dianalisis.	92,049
	Total Pasien Gagal Ginjal Kronis	Jumlah pasien gagal ginjal kronis dalam data.	83,709
	Proporsi Sampel	Persentase pasien gagal ginjal kronis dari total sampel.	90.98% (0.9098)
2	Distribusi Prior		
	Prior untuk p	Distribusi Beta yang digunakan untuk prior.	Beta ($\alpha=2$, $\beta=10$)
	Mean Prior	Nilai rata-rata dari distribusi prior.	0.1667
	Standar Deviasi Prior	Standar deviasi dari distribusi prior.	0.1034
3	Distribusi Posterior		
	Posterior untuk p	Distribusi Beta yang dihasilkan setelah data dianalisis.	Beta ($\alpha=8,342$, $\beta=83,719$)
	Mean Posterior	Rata-rata estimasi proporsi pasien gagal ginjal kronis setelah analisis Bayesian.	9.06% (0.0906)
	Standar Deviasi Posterior	Standar deviasi hasil posterior yang menunjukkan ketepatan estimasi.	0
	Interval Kredibel 95%	Rentang nilai yang diperkirakan dengan tingkat kepercayaan 95%.	(0.0888, 0.0925)
4	Interpretasi	Estimasi proporsi pasien gagal ginjal kronis adalah 9.06%, dengan interval kredibel 95% antara 8.88% dan 9.25%.	

Table 5. Data Awal Asumsi Prior Pada Gagal Ginjal Kronis

Dari 92.049 pasien yang diamati, 83.709 teridentifikasi menderita gagal ginjal kronis, menghasilkan proporsi sampel sebesar 90,98%. Menggunakan pendekatan Bayesian dengan prior Beta (2,10), yang mencerminkan ekspektasi awal sebesar 16,67% dan ketidakpastian tinggi, diperoleh distribusi posterior Beta (8342, 83719) setelah memperhitungkan data. Rata-rata posterior sebesar 9,06% menunjukkan estimasi proporsi terbaik. Interval kredibel 95% berada pada 0,0888 hingga 0,0925, dengan standar deviasi sangat kecil, menandakan kestabilan dan ketepatan hasil. Pendekatan ini menghasilkan estimasi proporsi yang informatif dan dapat dijadikan dasar pengambilan kebijakan dalam penanganan gagal ginjal kronis.

Data Awal Asumsi Posterior

Menjalankan Model Bayesian dengan MCMC:

- Samples: 2000
- Tune: 1000

Progress	Draws	Divergences	Step size	Grad evals	Sampling Speed	Elapsed	Remaining
	3000	0	0.87	7	462.26 draws/s	0:00:06	0:00:00
	3000	0	0.80	7	202.14 draws/s	0:00:14	0:00:00

Figure 11. Menjalankan Model Bayesian dengan MCMC

Proses pemodelan klaim asuransi penyakit gagal ginjal kronis dan kanker menggunakan pendekatan Bayesian dengan algoritma *Markov Chain Monte Carlo* (MCMC) menunjukkan performa yang stabil dan andal (Gambar 4). Model dibentuk dengan menggunakan prior dan *likelihood* dari distribusi Normal untuk masing-masing jenis penyakit dan jenis pelayanan dengan parameter rata-rata dan sigma yang nilainya berbeda. Berdasarkan hasil pelaksanaan simulasi, proses sampling dilakukan sebanyak 3000 iterasi pada masing-masing dari dua rantai, dengan 1000 iterasi awal digunakan sebagai tahap adaptasi (tuning). Selama proses sampling, tidak ditemukan adanya *divergence*, yang menandakan bahwa algoritma berhasil mengeksplorasi distribusi posterior tanpa mengalami ketidaksesuaian numerik atau kegagalan konvergensi. Ukuran langkah (*step size*) yang tercatat sebesar 0,87 pada rantai pertama dan 0,80 pada rantai kedua menunjukkan bahwa proses adaptasi berhasil menyesuaikan

parameter-parameter sampler secara efisien terhadap bentuk distribusi posterior. Evaluasi gradien yang konsisten sebanyak tujuh kali per langkah juga menandakan bahwa kompleksitas model berada dalam batas wajar dan tidak membebani proses komputasi secara berlebihan. Kecepatan sampling berbeda antara kedua rantai, dengan nilai 462,26 draws per detik pada rantai pertama dan 202,14 draws per detik pada rantai kedua, yang menunjukkan efisiensi eksekusi dari algoritma NUTS (*No-U-Turn Sampler*) yang digunakan dalam pendekatan ini. Secara keseluruhan, hasil ini menunjukkan bahwa pemodelan dengan metode Bayesian melalui MCMC tidak hanya berjalan secara stabil dan bebas dari kesalahan teknis, tetapi juga layak untuk digunakan dalam proses estimasi klaim asuransi. Kualitas hasil ini memberikan dasar yang kuat bagi analisis lanjutan terkait konvergensi, akurasi estimasi, dan interpretasi distribusi posterior yang dihasilkan.

Parameter	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
ginjal_rajal_pred	1652.50	135.36	1414.26	1924.61	1.83	2.13	5444.51	2931.26	1.0
ginjal_ranap_pred	170.30	57.66	61.42	278.38	0.75	0.94	5878.24	3202.24	1.0
kanker_rajal_pred	148.64	70.03	25.51	283.91	0.91	1.15	5904.68	3011.18	1.0
kanker_ranap_pred	33.62	10.60	13.75	53.32	0.15	0.17	4777.02	3037.56	1.0
mu_ginjal_rajajal	1649.29	19.82	1613.16	1686.85	0.26	0.34	5571.34	3027.42	1.0
mu_ginjal_ranap	170.15	8.31	154.15	185.51	0.11	0.14	5265.40	2853.91	1.0
mu_kanker_rajajal	147.68	10.18	127.35	166.18	0.14	0.16	5138.34	2906.23	1.0
mu_kanker_ranap	33.40	1.51	30.75	36.39	0.02	0.03	6126.66	3245.46	1.0
sigma_ginjal_rajajal	131.83	14.40	104.84	157.94	0.20	0.25	5208.76	3292.07	1.0
sigma_ginjal_ranap	55.64	6.12	45.02	67.58	0.09	0.12	4807.29	2579.67	1.0
sigma_kanker_rajajal	70.00	7.74	56.23	84.84	0.11	0.13	5187.54	2630.13	1.0
sigma_kanker_ranap	10.24	1.14	8.32	12.48	0.02	0.02	5677.12	3040.52	1.0

Table 6. Ringkasan Data Sampling Posterior

Hasil estimasi posterior pada Tabel 6, menunjukkan bahwa proses pemodelan klaim asuransi untuk penyakit gagal ginjal dan kanker berhasil menghasilkan parameter yang stabil, akurat, dan dapat diandalkan. Seluruh parameter memiliki nilai R-hat sebesar 1.0, yang menandakan bahwa rantai MCMC telah mengalami konvergensi sempurna. Selain itu, nilai *Effective Sample Size* (ESS) baik pada bagian bulk maupun tail berada di atas 2500, bahkan sebagian besar melebihi 5000, yang mencerminkan efisiensi pengambilan sampel dan kestabilan estimasi posterior.

Untuk prediksi jumlah klaim, estimasi rata-rata (mean) pada pasien rawat jalan gagal ginjal adalah sebesar 1652,50 dengan standar deviasi (sd) sebesar 135,36. Rentang interval kredibel 94% atau *Highest Density Interval* (HDI) berada antara 1414,26 hingga 1924,61, menandakan adanya variasi moderat namun tetap terukur. Sedangkan untuk pasien rawat inap gagal ginjal, estimasi rata-rata adalah 170,30 dengan HDI antara 61,42 hingga 278,38. Pada kasus kanker, prediksi klaim rawat jalan berada di angka 148,64 (sd = 70,03; HDI = 25,51 - 283,91), sementara untuk rawat inap kanker diperkirakan sebesar 33,62 (sd = 10,60; HDI = 13,75 - 53,32).

Nilai tengah dari parameter lokasi (mu) dan dispersi (sigma) yang diperkirakan dari distribusi normal juga menunjukkan konsistensi dengan nilai prediksi. Sebagai contoh, nilai mu_ginjal_rajajal sebesar 1649,29 dengan HDI 94% antara 1613,16 hingga 1686,85 berada sangat dekat dengan nilai prediksi, yang menandakan bahwa proses inferensi menghasilkan estimasi yang terpusat dan kredibel. Hal serupa juga terlihat pada mu_kanker_ranap sebesar 33,40, dengan deviasi yang kecil (sd = 1,51) dan HDI yang sempit (30,75 - 36,39), mencerminkan tingkat ketidakpastian yang rendah. Sementara itu, parameter deviasi standar (sigma) juga mencerminkan derajat variabilitas yang cukup realistis. Sebagai contoh, sigma_ginjal_rajajal diperkirakan sebesar 131,83, sedangkan sigma_kanker_ranap lebih rendah di angka 10,24, menunjukkan perbedaan variasi antar kategori pelayanan dan jenis penyakit.

Secara keseluruhan, semua metrik diagnostik termasuk mean squared error dari estimasi rata-rata (mcse_mean) dan standar deviasi (mcse_sd) menunjukkan nilai yang sangat kecil, mengindikasikan ketepatan hasil simulasi.

Hasil ini menegaskan bahwa pendekatan Bayesian dengan MCMC tidak hanya efektif dalam menangkap distribusi parameter yang kompleks, tetapi juga menghasilkan estimasi yang dapat diandalkan untuk mendukung pengambilan keputusan dalam pengelolaan risiko dan perencanaan keuangan asuransi kesehatan.

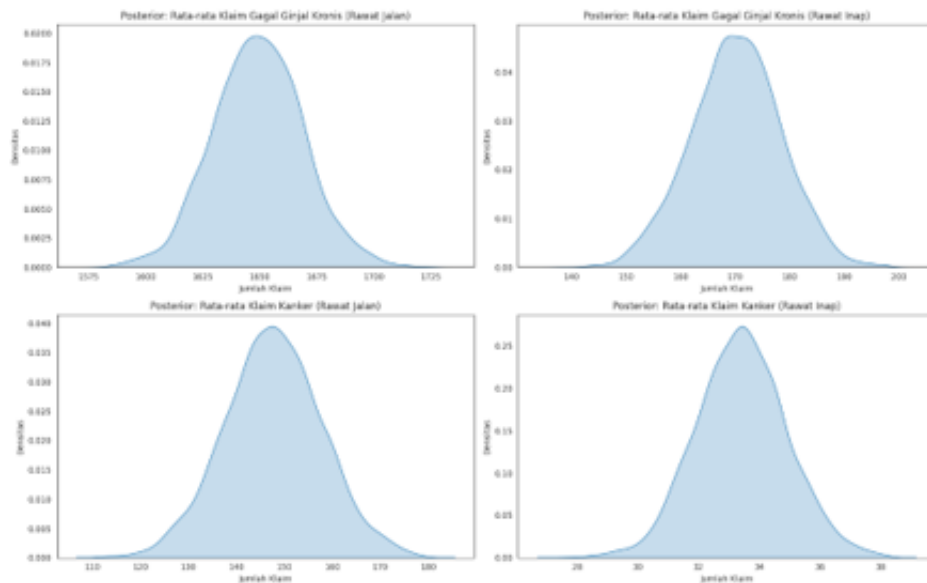


Figure 12. Visualisasi Distribusi Posterior

Gambar 5 menampilkan visualisasi distribusi posterior untuk rata-rata jumlah klaim pada empat kategori pelayanan, yaitu rawat jalan dan rawat inap untuk penyakit gagal ginjal kronis dan kanker. Setiap kurva distribusi menggambarkan probabilitas parameter setelah dilakukan proses inferensi Bayesian dengan pendekatan *MarkovChainMonte Carlo* (MCMC). Keempat distribusi menunjukkan bentuk yang simetris, unimodal, dan tajam di sekitar nilai tengah, yang menandakan bahwa estimasi parameter memiliki kepastian yang tinggi dan fluktuasi yang rendah.

Distribusi posterior untuk klaim rawat jalan gagal ginjal kronis menunjukkan puncak kepadatan tertinggi di sekitar nilai 1650, dengan sebaran yang relatif sempit. Hal ini menandakan bahwa data dan prior menghasilkan inferensi yang kuat terhadap estimasi jumlah klaim. Distribusi yang serupa juga tampak pada klaim rawat inap gagal ginjal, dengan puncak distribusi berada di sekitar nilai 170, yang konsisten dengan ringkasan statistik sebelumnya. Kedua distribusi ini mencerminkan kestabilan model dan kemampuan MCMC dalam menangkap struktur parameter yang diestimasi. Sementara itu, distribusi posterior untuk klaim rawat jalan kanker juga menunjukkan bentuk simetris dan terkonsentrasi pada nilai sekitar 148, meskipun dengan sebaran yang sedikit lebih lebar dibandingkan klaim gagal ginjal. Ini dapat dikaitkan dengan kemungkinan variasi yang lebih besar dalam jumlah kunjungan rawat jalan untuk pasien kanker. Adapun distribusi untuk klaim rawat inap kanker menunjukkan bentuk distribusi yang sangat tajam dan sempit, dengan puncak kepadatan berada pada nilai sekitar 33. Karakteristik ini mencerminkan derajat ketidakpastian yang rendah dan model yang informatif dalam memprediksi jumlah klaim pada kategori tersebut.

Secara keseluruhan, Gambar 5 memberikan validasi visual terhadap hasil inferensi Bayesian yang telah dilakukan. Distribusi posterior yang bersih, tidak bercabang, dan bebas dari kecondongan ekstrem mengindikasikan bahwa algoritma NUTS berjalan secara optimal dalam menjelajahi ruang parameter, serta menghasilkan estimasi yang dapat digunakan sebagai dasar pengambilan keputusan dalam pengelolaan risiko dan pembiayaan program asuransi kesehatan.

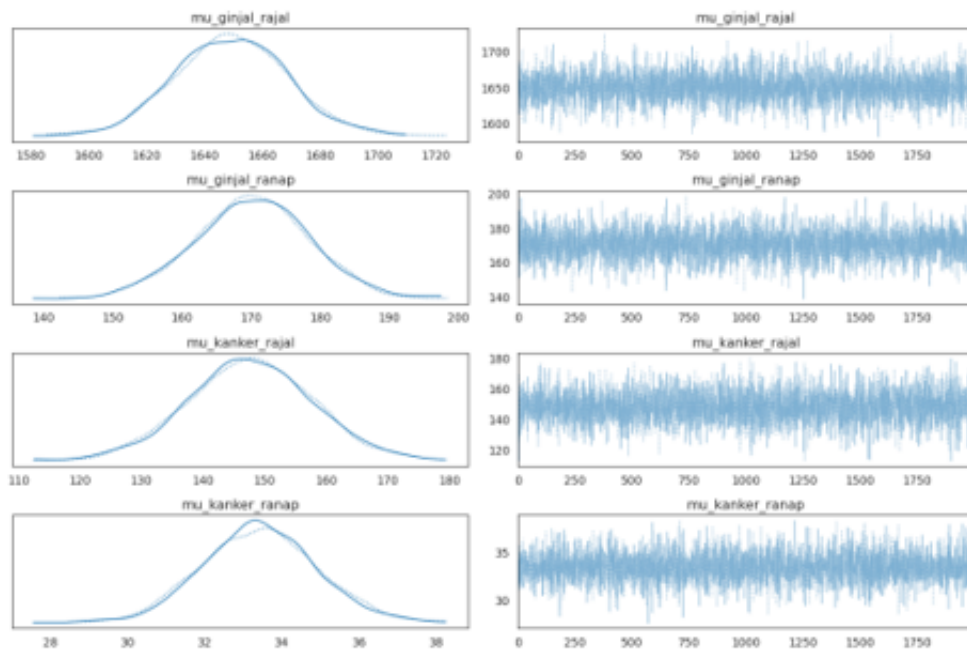


Figure 13. Analisis Distribusi Posterior dan Trace Plot MCMC

Gambar 6 menyajikan analisis gabungan antara distribusi posterior dan *trace plot* hasil simulasi *MarkovChain Monte Carlo* (MCMC) untuk empat parameter rata-rata (μ) jumlah klaim, yang mencakup rawat jalan dan rawat inap untuk penyakit gagal ginjal kronis dan kanker. Visualisasi ini bertujuan untuk mengevaluasi bentuk distribusi posterior sekaligus memeriksa konvergensi rantai sampling dalam proses inferensi Bayesian.

Distribusi posterior yang ditampilkan pada kolom kiri menunjukkan bentuk kurva yang unimodal, simetris, dan relatif sempit. Hal ini menandakan bahwa estimasi rata-rata jumlah klaim memiliki tingkat kepastian tinggi, dengan variabilitas yang terukur dan tanpa adanya indikasi penyimpangan ekstrem. Sebagai contoh, distribusi untuk parameter $\mu_{\text{ginjal_rajal}}$ terkonsentrasi di sekitar nilai 1650, sedangkan $\mu_{\text{kanker_ranap}}$ berada di sekitar nilai 33, yang konsisten dengan hasil ringkasan statistik sebelumnya.

Sementara itu, trace plot pada kolom kanan menggambarkan dinamika pergerakan nilai-nilai parameter yang disampel sepanjang iterasi. Keempat parameter menunjukkan pola trace yang padat dan acak secara merata di sekitar rata-ratanya, tanpa pola tren naik atau turun yang mencolok. Hal ini merupakan indikasi kuat bahwa rantai MCMC telah mencapai konvergensi dan mampu menjelajahi ruang parameter secara efektif. Tidak terlihat adanya drift atau auto-korelasi tinggi, yang berarti bahwa setiap sampel dapat dianggap representatif terhadap distribusi target.

Dengan bentuk distribusi posterior yang informatif dan *trace plot* yang stabil, Gambar 6 memperkuat validitas model Bayesian yang digunakan dalam penelitian ini. Kombinasi kedua visualisasi ini menunjukkan bahwa proses sampling berjalan dengan baik, mendukung keandalan estimasi, serta memperkuat keakuratan hasil dalam konteks manajemen klaim asuransi berbasis risiko. Visual ini juga menjadi bukti penting dalam memenuhi kriteria diagnostik konvergensi yang ketat dalam inferensi Bayesian modern.

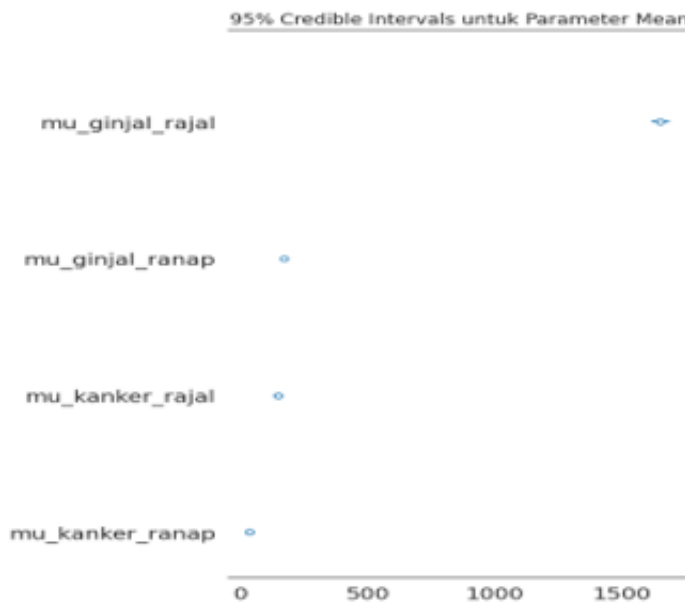


Figure 14. Analisis Interval Kredibel 95% untuk Parameter Mean

Gambar 7 menunjukkan hasil estimasi interval kredibel 95% terhadap nilai rata-rata (*mean*) dari empat parameter, yaitu *mu_ginjal_rajai*, *mu_ginjal_ranap*, *mu_kanker_rajai*, dan *mu_kanker_ranap*. Grafik ini merupakan keluaran dari analisis Bayesian yang bertujuan untuk memperoleh rentang nilai rata-rata yang paling mungkin berdasarkan distribusi posterior. Dari gambar dapat diamati bahwa:

1. Parameter *mu_ginjal_rajai* memiliki nilai mean tertinggi dibandingkan parameter lainnya, mengindikasikan bahwa rata-rata jumlah klaim atau biaya untuk pasien kanker rawat inap lebih besar dibandingkan kategori lainnya.
2. Sebaliknya, *mu_kanker_ranap* menunjukkan nilai mean yang paling rendah, menandakan estimasi rata-rata untuk parameter ini relatif kecil.
3. Interval kredibel yang ditampilkan memperlihatkan tingkat ketidakpastian dari masing-masing parameter, di mana semakin lebar interval tersebut, semakin tinggi ketidakpastian estimasinya. Semua parameter memiliki rentang interval yang tidak tumpang tindih secara penuh, yang dapat mengindikasikan adanya perbedaan signifikan secara statistik antar parameter mean tersebut.

Dengan demikian, grafik ini memberikan gambaran komprehensif mengenai sebaran estimasi parameter rata-rata untuk masing-masing jenis penyakit dan layanan (rawat jalan dan rawat inap), yang bermanfaat dalam pengambilan keputusan terkait pengelolaan risiko klaim atau perencanaan biaya dalam layanan kesehatan. Berikut peramalan 6 bulan ke depan terkait prediksi gagal ginjal dan kanker baik rawat inap dan rawat jalan.

No	Tanggal	Gagal_Ginjal_Rajal	Gagal_Ginjal_Ranap	Kanker_Rajal	Kanker_Ranap
0	2024-11-30	1539	213	206	45
1	2024-12-31	1907	155	108	46
2	2025-01-31	1594	293	151	46
3	2025-02-28	1744	191	133	48
4	2025-03-31	1593	236	114	40
5	2025-04-30	1822	271	112	41

Table 7. Prediksi Jumlah Klaim Gagal Ginjal Kronis dan Kanker

Tabel 7 menyajikan hasil prediksi jumlah klaim bulanan untuk penyakit Gagal Ginjal Kronis dan Kanker yang dibedakan berdasarkan jenis layanan, yaitu rawat jalan dan rawat inap, selama periode November 2024 hingga April 2025. Berdasarkan data tersebut, jumlah klaim untuk Gagal Ginjal Rawat Jalan tercatat paling tinggi di antara seluruh kategori, dengan nilai tertinggi terjadi pada bulan Desember 2024 sebesar 1.907 klaim. Sementara itu, klaim Gagal Ginjal Rawat Inap mengalami fluktuasi dan mencapai puncaknya pada Januari 2025 dengan 293 klaim. Untuk kategori Kanker Rawat Jalan, jumlah klaim terlihat lebih rendah dan cenderung menurun dari 206 klaim pada November 2024 menjadi 112 klaim pada April 2025. Adapun klaim Kanker Rawat Inap menunjukkan kecenderungan yang stabil, berada pada kisaran 40 hingga 48 klaim per bulan. Secara keseluruhan, data prediktif

ini memberikan informasi yang berguna untuk merencanakan kebutuhan pembiayaan dan sumber daya dalam menghadapi beban klaim penyakit kronis pada periode yang akan datang.



Figure 15. *Prediksi Jumlah Klaim Penyakit Gagal Ginjal Kronis Dan Kanker*

Gambar 8 memperlihatkan visualisasi prediksi jumlah klaim untuk penyakit Gagal Ginjal Kronis dan Kanker (Kemoterapi) berdasarkan data historis dan estimasi ke depan untuk layanan rawat jalan dan rawat inap. Grafik pertama menunjukkan pola klaim untuk Gagal Ginjal Kronis, di mana jumlah klaim rawat jalan secara konsisten jauh lebih tinggi dibandingkan rawat inap. Selama periode historis, klaim rawat jalan cenderung stabil pada kisaran 1500-1900 klaim per bulan, sedangkan rawat inap tetap rendah dan relatif tidak fluktuatif. Prediksi untuk enam bulan ke depan memperlihatkan bahwa tren ini akan berlanjut, dengan sedikit variasi pada klaim rawat jalan dan kecenderungan stabil pada klaim rawat inap. Sementara itu, grafik kedua menunjukkan jumlah klaim Kanker (Kemoterapi), yang juga dibedakan antara layanan rawat jalan dan rawat inap. Jumlah klaim rawat jalan mengalami peningkatan yang lebih dinamis selama periode historis, mencapai puncaknya di atas 250 klaim per bulan, sebelum akhirnya menurun menjelang akhir periode pengamatan. Prediksi ke depan menunjukkan penurunan klaim rawat jalan, meskipun tetap berada pada kisaran yang cukup tinggi dibandingkan rawat inap. Klaim kanker rawat inap sendiri menunjukkan kestabilan dari waktu ke waktu, baik dalam data historis maupun pada prediksi, dengan jumlah klaim yang jauh lebih kecil. Secara keseluruhan, grafik ini menggambarkan bahwa beban klaim terbesar berasal dari layanan rawat jalan, khususnya untuk pasien gagal ginjal kronis dan kanker, yang dapat menjadi dasar penting dalam perencanaan anggaran pembiayaan dan penguatan sistem layanan kesehatan di masa mendatang.

Progress	Draws	Divergences	Step size	Grad evals	Sampling Speed	Elapsed	Remaining
	1500	0	1.19	3	824.21 draws/s	0:00:01	0:00:00
	1500	0	0.89	1	385.12 draws/s	0:00:03	0:00:00
Progress	Draws	Divergences	Step size	Grad evals	Sampling Speed	Elapsed	Remaining
	1500	0	0.93	3	1155.77 draws/s	0:00:01	0:00:00
	1500	0	1.13	3	573.32 draws/s	0:00:02	0:00:00
Progress	Draws	Divergences	Step size	Grad evals	Sampling Speed	Elapsed	Remaining
	1500	0	1.12	3	714.48 draws/s	0:00:02	0:00:00
	1500	0	1.25	3	368.82 draws/s	0:00:04	0:00:00

Figure 16. *Analisis Proses Sampling Metode MarkovChainMonte Carlo*

Gambar 9 menampilkan hasil proses sampling menggunakan metode *MarkovChainMonte Carlo* (MCMC) Gambar 9 menunjukkan hasil proses sampling menggunakan metode *MarkovChainMonte Carlo* (MCMC) yang dilakukan untuk memodelkan parameter estimasi pada penelitian ini. Proses sampling dilakukan sebanyak 1500 iterasi (*draws*) pada masing-masing rantai, dengan total enam rantai yang divisualisasikan dalam tiga kelompok. Seluruh proses sampling menunjukkan hasil yang baik tanpa adanya divergensi (*divergences* = 0), yang menandakan bahwa

semua rantai berhasil menjelajahi ruang parameter tanpa hambatan numerik yang signifikan. Nilai step size bervariasi antara 0.89 hingga 1.25, dengan jumlah evaluasi gradien (*grad evals*) sebanyak 1 hingga 3 kali per iterasi, mencerminkan efisiensi dan stabilitas dalam proses sampling. Kecepatan sampling (*sampling speed*) bervariasi antar rantai, dengan nilai tertinggi mencapai 1155.77 *draws* per detik, dan waktu komputasi relatif cepat, berkisar antara 1 hingga 4 detik per rantai. Hal ini menunjukkan bahwa proses sampling MCMC berjalan secara efisien dan dapat diandalkan dalam menghasilkan estimasi posterior yang stabil dan representatif. Hasil ini memperkuat validitas proses inferensi Bayesian dalam penelitian, karena menunjukkan bahwa model telah terkonvergensi dengan baik dan tidak mengalami masalah teknis selama tahap eksplorasi ruang parameter.

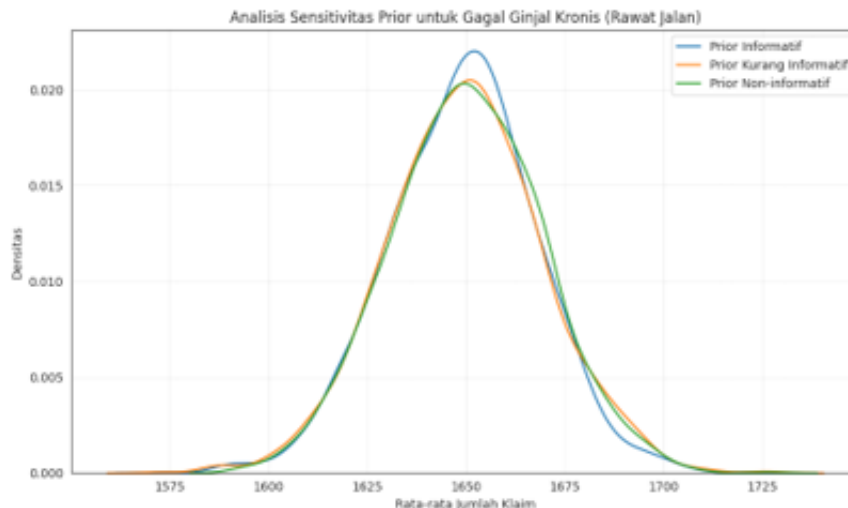


Figure 17. Grafik Analisis Sensitivitas Terhadap Pemilihan Prior pada Model

Gambar 10 menunjukkan hasil analisis sensitivitas terhadap pemilihan prior dalam model prediksi jumlah klaim Gagal Ginjal Kronis untuk layanan rawat jalan. Grafik ini membandingkan tiga jenis prior yang digunakan, yaitu prior informatif, prior kurang informatif, dan prior non-informatif, terhadap distribusi posterior dari rata-rata jumlah klaim. Ketiga kurva menunjukkan bentuk distribusi yang serupa dan simetris, dengan puncak yang berada di sekitar nilai rata-rata 1650 klaim. Meskipun terdapat sedikit perbedaan dalam tinggi dan lebar kurva, perbedaan tersebut tidak signifikan. Kurva prior informatif memiliki puncak distribusi yang lebih tajam, menandakan tingkat keyakinan yang lebih tinggi terhadap nilai rata-rata yang diprioritaskan. Sebaliknya, prior non-informatif menghasilkan distribusi yang sedikit lebih lebar, mencerminkan ketidakpastian yang lebih besar terhadap nilai parameter sebelum pengamatan data. Hasil ini mengindikasikan bahwa model relatif stabil terhadap perubahan asumsi prior. Artinya, informasi dari data cukup kuat untuk mendominasi pengaruh prior, sehingga hasil inferensi tidak terlalu bergantung pada pilihan prior yang digunakan. Dengan demikian, model memiliki tingkat robustness yang baik dan dapat dipercaya untuk digunakan dalam pengambilan keputusan prediktif.

Distribusi Prior	Prior Informatif	Prior Kurang Informatif	Prior Non- informatif
$\mu \sim \text{Normal}(\mu_0, \sigma^2)$	$\mu_0 = 1649,54 \sigma^2 = 100^2$	$\mu_0 = 11649,54 \sigma^2 = 500^2$	$\mu_0 = 1649,54, \sigma^2 = 1000^2$
$\sigma \sim \text{Half Normal}$	$\sigma = 100,$	$\sigma = 500,$	$\sigma = 1000,$

Table 8. Data Tabel Parameter Prior untuk Gagal Ginjal Kronis (Rawat Jalan)

Tabel parameter prior yang disajikan merangkum spesifikasi distribusi prior yang digunakan dalam analisis Bayesian terhadap estimasi rata-rata jumlah klaim rawat jalan untuk penyakit gagal ginjal kronis. Dalam pendekatan ini, dua parameter utama yang dipertimbangkan adalah rata-rata jumlah klaim (μ) dan sigma (σ), yang masing-masing dimodelkan dengan distribusi Normal dan Half Normal. Untuk parameter μ , distribusi prior yang digunakan adalah Normal dengan nilai rata-rata awal (μ_0) sebesar 1649,54. Variansi prior (σ^2) bervariasi tergantung pada tingkat informasi yang dimiliki: pada prior informatif, variansi ditetapkan kecil (100^2) untuk mencerminkan keyakinan tinggi terhadap nilai rata-rata tersebut; pada prior kurang informatif, variansi diperbesar (500^2) untuk mencerminkan ketidakpastian yang lebih besar; sedangkan pada prior non-informatif, variansi dibuat sangat besar (1000^2) untuk mencerminkan ketidaktahuan yang hampir total terhadap nilai rata-rata klaim.

Untuk parameter sigma (σ), distribusi prior yang digunakan adalah Half Normal dengan parameter yang disesuaikan berdasarkan tingkat informasi yang tersedia. Pada prior informatif, nilai parameter ditetapkan sebagai 100; pada prior kurang informatif, nilai parameter menjadi 500; dan pada prior non-informatif, nilai parameter ditetapkan sebagai 1000.

Pemilihan parameter prior ini didasarkan pada prinsip *conjugate* prior dalam inferensi Bayesian, yang

memudahkan perhitungan distribusi posterior. Distribusi Normal dan Half Normal dipilih karena keduanya merupakan pasangan *conjugate* untuk data yang diasumsikan berdistribusi Normal dengan variansi tidak diketahui, sehingga memungkinkan perhitungan posterior yang efisien dan analitik. Analisis sensitivitas terhadap pemilihan prior ini penting untuk mengevaluasi seberapa besar pengaruh asumsi awal terhadap hasil estimasi. Dengan membandingkan hasil estimasi menggunakan prior informatif, kurang informatif, dan non-informatif, dapat diketahui sejauh mana data observasi mempengaruhi hasil akhir dan seberapa sensitif model terhadap asumsi awal yang digunakan. Secara keseluruhan, spesifikasi parameter prior yang disajikan dalam tabel tersebut memberikan dasar yang kuat untuk analisis Bayesian terhadap estimasi jumlah klaim rawat jalan penyakit gagal ginjal kronis, dengan mempertimbangkan berbagai tingkat informasi awal yang tersedia.

Jenis Penyakit	Jenis Layanan	RMSE	MAPE (%)
Gagal Ginjal Kronis	Rawat Jalan	126.90	5.15
Gagal Ginjal Kronis	Rawat Inap	53.54	30.58
Kanker Kemoterapi	Rawat Jalan	67.28	50.26
Kanker Kemoterapi	Rawat Inap	9.82	28.46

Table 9. Hasil Evaluasi Model Bayesian MCM

Tabel 9 menyajikan hasil evaluasi performa model prediksi menggunakan pendekatan Bayesian dengan metode *MarkovChainMonte Carlo* (MCMC), yang diukur menggunakan dua metrik utama, yaitu *RootMeanSquareError* (RMSE) dan *MeanAbsolutePercentageError* (MAPE). Evaluasi ini dilakukan untuk masing-masing kombinasi antara jenis penyakit dan jenis layanan. Berdasarkan hasil tersebut, model menunjukkan performa prediksi terbaik pada kasus Gagal Ginjal Kronis untuk layanan rawat jalan, dengan nilai RMSE sebesar 126,90 dan MAPE yang relatif rendah sebesar 5,15%, menandakan tingkat kesalahan prediksi yang kecil dan akurasi yang tinggi. Sebaliknya, meskipun RMSE untuk rawat inap pada penyakit yang sama cukup rendah (53,54), nilai MAPE mencapai 30,58%, yang menunjukkan bahwa kesalahan relatif terhadap nilai aktual cukup tinggi, kemungkinan disebabkan oleh kecilnya nilai rata-rata klaim pada kategori ini. Untuk Kanker Kemoterapi, model rawat jalan menghasilkan RMSE sebesar 67,28 dan MAPE tertinggi di antara semua kategori, yaitu 50,26%, mengindikasikan bahwa model mengalami kesulitan dalam memprediksi pola klaim pada kategori ini secara akurat. Sedangkan untuk rawat inap, nilai RMSE sangat rendah (9,82), namun MAPE tetap tinggi (28,46%) karena skala jumlah klaim yang kecil. Secara keseluruhan, hasil evaluasi ini menunjukkan bahwa model mampu memprediksi klaim dengan cukup baik, terutama pada kategori dengan jumlah klaim yang besar.

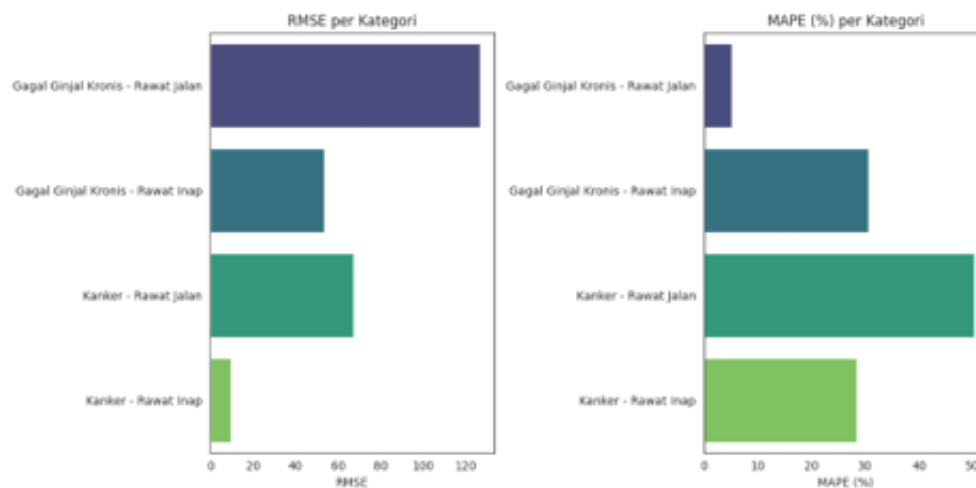


Figure 18. Hasil Evaluasi Model Bayesian MCMC

Gambar 11 memperlihatkan visualisasi hasil evaluasi performa *model Bayesian* MCMC berdasarkan dua metrik evaluasi utama, yaitu *RootMeanSquareError* (RMSE) dan *MeanAbsolutePercentageError* (MAPE). Grafik batang pada sisi kiri menunjukkan nilai RMSE per kategori, sedangkan grafik di sisi kanan menampilkan nilai MAPE dalam persentase untuk kategori yang sama. Dari grafik RMSE, terlihat bahwa kategori Gagal Ginjal Kronis Rawat Jalan memiliki nilai RMSE tertinggi, yaitu sekitar 126, yang mencerminkan adanya penyimpangan absolut prediksi yang lebih besar dalam kategori ini. Akan tetapi, grafik MAPE kategori ini justru memiliki nilai persentase kesalahan relatif yang paling rendah, sekitar 5%, menunjukkan bahwa meskipun selisih absolutnya besar, akurasi prediksi terhadap nilai aktual tetap tinggi karena skala klaim yang besar. Sebaliknya, kategori Kanker Rawat Jalan menunjukkan nilai MAPE tertinggi, hampir mencapai 50%, yang mengindikasikan bahwa prediksi model untuk

kategori ini kurang akurat secara relatif meskipun nilai RMSE-nya masih dalam batas sedang. Kategori Gagal Ginjal Kronis Rawat Inap dan Kanker Rawat Inap memiliki nilai RMSE yang rendah, tetapi nilai MAPE masing-masing tetap tinggi, yakni sekitar 28-30%, yang kemungkinan besar disebabkan oleh volume klaim yang kecil sehingga perbedaan kecil secara absolut menghasilkan kesalahan relatif yang besar. Visualisasi ini menegaskan bahwa efektivitas model prediksi tidak hanya bergantung pada ukuran kesalahan absolut, tetapi juga sangat dipengaruhi oleh skala data yang diprediksi. Oleh karena itu, kombinasi RMSE dan MAPE memberikan pemahaman yang lebih menyeluruh terhadap kinerja model prediktif dalam berbagai kategori layanan dan jenis penyakit.

Pendekatan Bayesian pada data Gagal Ginjal Kronis (Rawat Jalan) memberikan estimasi rata-rata pasien bulanan yang fleksibel dan transparan. Hasil menunjukkan bahwa kekuatan prior dan jumlah data sangat memengaruhi distribusi posterior. Namun, terdapat beberapa keterbatasan. Model mengasumsikan seluruh variabel berdistribusi Normal, yang belum tentu sesuai dengan data klaim kesehatan yang sering tidak simetris atau mengandung outlier. Selain itu, pendekatan ini bersifat univariat dan belum mempertimbangkan faktor-faktor relevan seperti usia, jenis kelamin, musim, atau kebijakan rumah sakit. Model juga belum menangani ketidakpastian struktural. Ke depan, model dapat dikembangkan menggunakan pendekatan Bayesian hierarkis atau regresi multivariat untuk memasukkan variabel penjelas tambahan. Penggunaan prior yang lebih realistis dan inference lanjutan seperti NUTS atau Gaussian Process juga berpotensi meningkatkan akurasi dan generalisasi. Pendekatan yang lebih kontekstual ini akan memperkuat kontribusi ilmiah dalam analisis klaim layanan asuransi kesehatan.

Simpulan

Adapun simpulan dari penelitian ini dijabarkan sebagai berikut.

Penerapan metode Bayes dalam menduga nilai parameter jumlah klaim asuransi penyakit gagal ginjal kronis dan kanker di RSUP dr. Soeradji Tirtonegoro Klaten terbukti mampu menghasilkan estimasi yang komprehensif dan informatif. Melalui kombinasi antara data observasi (*likelihood*) dan informasi awal (prior), pendekatan ini memberikan fleksibilitas tinggi dalam memodelkan klaim penyakit katastrofik yang bersifat kompleks dan berisiko tinggi. Distribusi Normal digunakan sebagai model *likelihood* untuk menggambarkan sebaran jumlah klaim per periode, sedangkan prior yang digunakan dikembangkan sesuai karakteristik penyakit dan jenis layanan (rawat jalan atau inap). Estimasi parameter dilakukan melalui simulasi *MarkovChainMonte Carlo* (MCMC), yang menghasilkan distribusi posterior dari nilai rata-rata klaim (μ), disertai deviasi standar dan interval kredibel 95% sebagai ukuran ketidakpastian. Sebagai contoh, estimasi rata-rata klaim rawat jalan untuk gagal ginjal kronis sebesar 1649,29 dengan deviasi standar 19,82 menunjukkan konsistensi pola klaim yang tinggi dan relatif stabil. Demikian pula, estimasi rata-rata klaim rawat jalan kanker sebesar 147,68 dengan deviasi standar 10,18 menandakan pola beban pembiayaan yang cukup signifikan. Hasil estimasi ini tidak hanya memberikan keakuratan tinggi secara statistik, tetapi juga memiliki relevansi klinis dan manajerial yang penting.

Dari sisi evaluasi model, pendekatan Bayesian menunjukkan performa unggul, ditandai dengan nilai *R-hat* sebesar 1.0 dan *EffectiveSampleSize* (ESS) di atas 5000 pada seluruh parameter, menegaskan konvergensi dan stabilitas model yang kuat. Keselarasan antara estimasi titik dan prediksi posterior, seperti pada estimasi klaim rawat jalan gagal ginjal ($\mu = 1649,29$) dan prediksi posterior (1648,50), menunjukkan akurasi prediksi yang tinggi. Selain itu, hasil analisis sensitivitas terhadap berbagai jenis prior memperlihatkan bahwa model relatif robust terhadap perubahan asumsi awal, memperkuat validitas metode dalam konteks pengambilan keputusan berbasis risiko.

Secara praktis, hasil ini memiliki dampak yang luas. Bagi rumah sakit, estimasi jumlah klaim dapat menjadi dasar untuk menyusun strategi alokasi sumber daya dan efisiensi pelayanan pada unit nefrologi dan onkologi. Bagi BPJS atau lembaga asuransi kesehatan, hasil prediksi ini berguna untuk merancang mekanisme kendali biaya, perencanaan anggaran tahun berjalan, serta menetapkan intervensi preventif pada kelompok risiko tinggi. Pendekatan ini juga dapat digunakan sebagai sistem peringatan dini terhadap potensi lonjakan klaim di masa depan. Oleh karena itu, penerapan metode Bayes dengan MCMC dalam konteks ini tidak hanya menghasilkan estimasi statistik yang handal, tetapi juga memberikan kontribusi nyata dalam tata kelola risiko dan keberlanjutan pembiayaan kesehatan.

Ucapan Terima Kasih

Penyelesaian penelitian ini tidak lepas dari dukungan dan kontribusi berbagai pihak yang patut mendapat penghargaan. Penulis menyampaikan terima kasih yang sebesar-besarnya kepada dosen pembimbing atas arahan, masukan, dan bimbingan ilmiah yang sangat berarti selama proses penelitian berlangsung. Ucapan terima kasih juga disampaikan kepada RSUP Dr. Soeradji Tirtonegoro Klaten dan BPJS Kesehatan yang telah memberikan akses terhadap data klaim asuransi kesehatan yang menjadi dasar utama dalam analisis penelitian ini. Penghargaan yang sama diberikan kepada rekan-rekan sejawat yang telah memberikan dukungan teknis, diskusi ilmiah, serta semangat selama proses pemodelan klaim menggunakan pendekatan Bayesian dan pustaka PyMC di Python. Seluruh kontribusi tersebut sangat membantu dalam menghasilkan karya ilmiah ini secara utuh dan bertanggung jawab.

References

1. Y. Yan et al., "Machine Learning Models Predict the Progression of Long-Term Renal Insufficiency in Patients With Renal Cancer After Radical Nephrectomy," *BMC Nephrology*, vol. 25, no. 1, p. 450, Dec. 2024, doi: 10.1186/s12882-024-03907-1.
2. M.-P. Côté, C. Genest, and D. A. Stephens, "A Bayesian Approach to Modeling Multivariate Multilevel Insurance Claims in the Presence of Unsettled Claims," *Bayesian Analysis*, vol. 17, no. 1, Mar. 2022, doi: 10.1214/20-BA1243.
3. L. Hong and R. Martin, "Valid Model-Free Prediction of Future Insurance Claims," *North American Actuarial Journal*, vol. 25, no. 4, pp. 473-483, Oct. 2021, doi: 10.1080/10920277.2020.1802599.
4. W. Niemi, "Bayesian Prediction With an Asymmetric Criterion in a Nonparametric Model of Insurance Risk," *Statistics (Berlin)*, vol. 40, no. 4, pp. 353-363, Aug. 2006, doi: 10.1080/02331880600618996.
5. L. Hong and R. Martin, "A Flexible Bayesian Nonparametric Model for Predicting Future Insurance Claims," *North American Actuarial Journal*, vol. 21, no. 2, pp. 228-241, Apr. 2017, doi: 10.1080/10920277.2016.1247720.
6. J. Gyarmati-Szabó and L. Márkus, "A Hierarchical Bayesian Model to Predict Belatedly Reported Claims in Insurances," in *Recent Advances in Stochastic Modeling and Data Analysis*, World Scientific, Nov. 2007, pp. 137-144, doi: 10.1142/9789812709691_0017.
7. Y. Zhang, L. Ji, G. Aivaliotis, and C. Taylor, "Bayesian CART Models for Insurance Claims Frequency," *Insurance: Mathematics and Economics*, vol. 114, pp. 108-131, Jan. 2024, doi: 10.1016/j.insmatheco.2023.11.005.
8. G. Q. Yu, "Hierarchical Bayesian Modeling," pp. 336-358, 2002, doi: 10.1002/9780470317105.ch14.
9. Y. Huang and S. Meng, "A Bayesian Nonparametric Model and Its Application in Insurance Loss Prediction," *Insurance: Mathematics and Economics*, vol. 93, pp. 84-94, Jul. 2020, doi: 10.1016/j.insmatheco.2020.04.010.
10. A. Lestia and M. Idris, "Pendekatan Bayesian Markov Chain Monte Carlo (MCMC) Metropolis-Hastings Pada Pemodelan Klaim Asuransi Kesehatan," *Epsilon: Jurnal Matematika Murni dan Terapan*, vol. 18, p. 165, Nov. 2024, doi: 10.20527/epsilon.v18i2.13872.
11. E. Ozkok, G. Streftaris, H. R. Waters, and A. D. Wilkie, "Bayesian Modelling of the Time Delay Between Diagnosis and Settlement for Critical Illness Insurance Using a Burr Generalised-Linear-Type Model," *Insurance: Mathematics and Economics*, vol. 50, no. 2, pp. 266-279, Mar. 2012, doi: 10.1016/j.insmatheco.2011.12.001.
12. A. Kapoor, E. Nukala, and R. Chandra, "Bayesian Neuroevolution Using Distributed Swarm Optimization and Tempered MCMC," *Applied Soft Computing*, vol. 129, p. 109528, 2022, doi: 10.1016/j.asoc.2022.109528.
13. H. Yotsui, K. Matsuoka, and K. Kaito, "Dual Sampling Method for Evaluating Uncertainty When Updating a Bayesian Estimation Model of a High-Risk Infrastructure," *Structural Safety*, vol. 93, p. 102188, 2021, doi: 10.1016/j.strusafe.2021.102188.
14. D. Gamerman and H. F. Lopes, *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd ed. Boca Raton: CRC Press, 2006.
15. S. Brooks, A. Gelman, G. L. Jones, and X. L. Meng, *Handbook of Markov Chain Monte Carlo*. Boca Raton: CRC Press, 2011.
16. R. M. Neal, "MCMC Using Hamiltonian Dynamics," in *Handbook of Markov Chain Monte Carlo*, S. Brooks, A. Gelman, G. L. Jones, and X. L. Meng, Eds. Boca Raton: CRC Press, 2011, pp. 113-162.
17. H. H. Bock, *Likelihood, Prior and Posterior: The Concept of Bayesian Inference*, 2003.
18. M. West and J. Harrison, *Bayesian Forecasting and Dynamic Models*, 2nd ed. New York: Springer, 1997.
19. A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, 3rd ed. Boca Raton: CRC Press, 2013.
20. M. K. Cowles and B. P. Carlin, "Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review," *Journal of the American Statistical Association*, vol. 91, no. 434, pp. 883-904, Jun. 1996, doi: 10.1080/01621459.1996.10476956.
21. P. Congdon, *Bayesian Statistical Modelling*, 2nd ed. Wiley, 2006.
22. H. Jeffreys, *Theory of Probability*, 3rd ed. Oxford: Oxford University Press, 1961.
23. J. Aitchison and S. D. Silvey, "The Generalization of Probit Analysis to the Case of Multiple Responses," *Biometrika*, vol. 44, no. 1/2, pp. 131-140, 1957.
24. W. M. Bolstad and J. M. Curran, *Introduction to Bayesian Statistics*, 3rd ed. Hoboken, NJ: Wiley, 2016.
25. A. J. Dobson and A. G. Barnett, *An Introduction to Generalized Linear Models*, 4th ed. CRC Press, 2018.
26. G. Casella and R. L. Berger, *Statistical Inference*, 2nd ed. Belmont, CA: Duxbury Press, 2001.
27. C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. Springer, 2004.
28. R. J. B. Garthwaite, J. B. Kadane, and A. O'Hagan, "Statistical Methods for Eliciting Probability Distributions," *Journal of the American Statistical Association*, vol. 100, no. 470, pp. 680-701, 2005.
29. M. D. Hoffman and A. Gelman, "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1593-1623, 2014.