Bayesian Changepoint Detection

# UNCERTAINTY ASSESSMENT & BAYESIAN COMPUTATION

Ilsin Su Sahin

# Introduction

## Background Information

Historically, coal mining has been dangerous, resulting in many accidents and a high death toll. The absence of safety regulations in the 19th century made this worse, as evidenced by the underground storage of explosives and the lack of ventilation requirements in mines. The United Kingdom's Coal Mines Regulations Act of 1887, which required ventilation and raised the minimum working age from 10 to 12 years old, was a crucial step in enhancing these conditions.

The inherent risks associated with underground mining as well as the absence of regulatory frameworks that could have reduced some of these risks contributed to the hazardous nature of coal mining, particularly in the 19th century. The working environment was appalling, with miners constantly at risk of explosions, collapses, and long-term health problems from inadequate ventilation. The coal mines regulations act of 1887 was only passed by the UK government in response to a number of tragedies and public outcry. This Act, which raised the legal working age to avoid child labour and introduced mandatory safety measures like proper ventilation to disperse deadly gases, was a historic reform. With the goal of shielding employees from the severe risks that had previously been the norm, these regulations signalled the start of a structural shift in the mining sector. Mining remained a hazardous occupation despite these developments, and more research into the frequency of disasters may be able to shed light on the success of the initial initiatives to increase mine safety as well as the ensuing development of mining techniques.

## Objectives of the Report

This report seeks to apply Bayesian changepoint detection to the coal mining disaster dataset to identify points in time where the underlying disaster rate may have changed. Detecting such changepoints is crucial for understanding shifts in the frequency of disasters, which may be associated with changes in safety regulations, mining practices, or external factors affecting the mining industry.

The primary objective is to model the disaster frequency data using Bayesian statistics and detect potential changepoints. The report aims to compare different models to assess the most probable changepoint locations and disaster rates before and after these changepoints. Furthermore, it endeavours to evaluate the robustness and scalability of the Bayesian methods used.

# Literature Review

## Introduction to Bayesian Changepoint Detection

A statistical method called "changepoint detection" is used to find locations in a time-ordered dataset where a change in the probability distribution occurs. This method is especially important for industries that need to accurately identify changes in data trends, like manufacturing quality control, EEG monitoring of brain activity, genetic sequence analysis, financial market analysis, and disease incidence rate tracking (Adams & MacKay, 2007). Changepoint detection has historically progressed from early frequentist techniques, which offered online detection capabilities, to more advanced Bayesian algorithms. By sampling from posterior distributions, Bayesian approaches are generally used in offline analysis to retroactively determine changepoints (Fearnhead, 2006). Online inference methods based on causal predictive filtering have been made possible by recent advances in Bayesian changepoint detection. In the fields of autonomous robotics and adaptive vision systems, where environmental changes must be quickly assimilated and responded to, these algorithms offer real-time, dynamic estimations of future data points based only on past data (Adams & MacKay, 2007). The vital relevance of Bayesian changepoint detection in contemporary data analysis is highlighted by its numerous applications in various scientific areas, which attest to its versatility and robustness.

## Metropolis-Hastings Sampler

One critical instrument in the instruments of Bayesian statistical approaches, especially in the context of Markov Chain Monte Carlo (MCMC) methods, is the Metropolis-Hastings (MH) algorithm. It is intended to sample from complex probability distributions, particularly in situations where direct sampling is not feasible because to the complexity of the distribution. This procedure, which is essential to Bayesian inference, creates a posterior distribution—which is frequently only known up to a multiplicative constant—by combining prior beliefs with the likelihood of observed data.

The MH algorithm initiates with an arbitrary starting point within the support of the target distribution. It then iterates through a sequence of steps where a candidate sample is drawn from a proposal distribution, and whether this sample is accepted as the next state in the Markov chain is determined by an acceptance probability. This probability hinges on the ratio of the product of the target distribution's density at the candidate sample and the proposal density transitioning to the current sample, compared to the product of the target distribution's density at the current sample and the proposal density transitioning to the candidate sample. The formula for the acceptance probability is expressed as:

$$\alpha(x' \mid x_t) = min\left(1, \frac{p(x')q(x_t \mid x')}{p(x_t)q(x' \mid x_t)}\right)$$

where p (·) denotes the target density and q (·|·) denotes the proposal density.

Because of this characteristic, the MH algorithm is especially useful for Bayesian inference problems when the normalisation constant of the posterior distribution is uncertain. But there are drawbacks to this flexibility as well. The proposal distribution must be carefully chosen to strike a balance between convergence to the target distribution and parameter space exploration in order for the algorithm to operate efficiently. The multimode distribution or high-dimensional problems might pose significant computational challenges when utilising the MH algorithm.

The MH algorithm is used to stochastically explore the parameter space for possible changepoints in the context of Bayesian changepoint detection, negotiating the complexities of mean and variance shifts. Notwithstanding these difficulties, the MH algorithm's ability to handle complicated distributions and its adaptability in use have cemented its position as an essential instrument for Bayesian inference, especially in the nuanced field of changepoint detection.

# Methodology

## Data Description

The dataset that is currently being examined is derived from Jarrett's 1979 research and is frequently used in the investigation of coal mining disaster frequencies. The National Coal Board in London released the Colliery Yearbook and Coal Trades Directory, which contained historical information used to build this dataset. The dataset offers a historical overview of UK coal mine explosions, covering the years 1851 through 1962, when the publishing came to a stop. Figure 1 displays the total data on coal mining accidents over the course of 111 years (Jarrett, 1979).

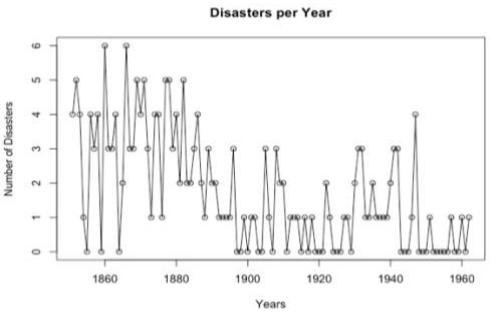


**Figure 1:** The plot illustrates the annual number of coal mining disasters in the United Kingdom from 1851 to 1962. A marked decrease in the frequency of disasters is observable following the late 1880s.

The annual number of disasters varies noticeably during the course of the timeframe. Certain years exhibit a high frequency of disasters; some years may have as many as six disasters. There seems to have been a decline in disasters following the 1880s, which is consistent with past accounts of enhanced mining safety legislation around that time. The frequency of disasters annually has generally decreased in the latter years, especially around 1900, which may indicate improvements in safety protocols or mining techniques during this period.

## Bayesian Framework

The annual disaster counts, represented by y, are treated as data from a Poisson process in the Bayesian paradigm for changepoint detection. The rate parameter λ, which describes this process, is thought to be changeable because of possible changepoints during the historical era. The model space extends to more sophisticated models with one or more changepoints, starting with a baseline model that assumes a constant rate λ over the whole-time span. We specify prior distributions for λ and any possible changepoint locations T before we observe the data. These priors capture our knowledge of the parameter values prior to the data. since of its conjugate qualities regarding the Poisson probability, a Gamma prior is typically chosen for λ since it makes the posterior computation easier.
Determining the posterior distribution of the parameters based on the data, $p(\lambda,T-y)$, is essential for determining the existence and positions of changepoints. The joint posterior distribution is difficult to approximate, especially when several changepoints are considered. For this reason, computational methods like Markov Chain Monte Carlo (MCMC) are used. The posterior distributions of λ and T are used to evaluate the model's predictive ability, and the marginal likelihood is used to examine how well the model fits the data. The model evidence is produced by integrating the joint posterior over all parameters to determine this likelihood. This Bayesian technique provides a comprehensive framework for the examination of changes in catastrophe rates over time, while also accommodating the inherent uncertainty in the changepoint detection procedure and allowing for the integration of past historical knowledge.

## Prior Selection and Justification

In order to include prior knowledge or assumptions about the parameters into the model, choosin g priors is a fundamental stage in the Bayesian analytical process.Priors for the Poisson process' s rate parameter λ and the changepoint locations T are carefully selected in the context of change point detection in order to take into account the historical background and the characteristics of coal mining accidents.Because of its conjugacy with the Poisson probability, which makes posteri or updating easier, a Gamma prior for λ is frequently justified.Assuming a priori that every point in time has an equal chance of becoming a changepoint, the choice of priors for T often follows a uniform distribution over the range of potential changepoint locations.
In our analysis, the rate parameter λ for the Poisson process was assigned a gamma prior distribution with a shape parameter of 2 and a rate of 1. The gamma distribution is a conjugate prior for the Poisson likelihood, facilitating analytical tractability. This prior reflects a belief in a relatively low but uncertain disaster rate, with the flexibility to adjust as data is observed.

## Inference Methods

The posterior distributions of the model parameters, in particular the locations and rates of the changepoints, are computed as part of the inference procedures for Bayesian changepoint detection. The work involves a lot of calculation, especially when there are several changepoints. For this reason, Markov Chain Monte Carlo (MCMC) techniques like the Metropolis-Hastings algorithm are used. Following a burn-in period, the samples produced by these iterative methods from the posterior distribution closely resemble the distribution. To analyse the MCMC output and draw conclusions about the changepoints, posterior summarization methods and convergence diagnostics are used.
In our analysis, Bayesian inference was conducted through the Metropolis-Hastings algorithm, a type of Markov Chain Monte Carlo (MCMC) method, to generate samples from the posterior

distribution of λ. This approach allows us to approximate the posterior when direct sampling is challenging, particularly in complex models or with large datasets.

## Convergence Diagnostics

Our Markov Chain Monte Carlo (MCMC) simulations' convergence must be thoroughly evaluated to guarantee the accuracy of our Bayesian estimations. To ensure that the chains have sufficiently explored the parameter space and that the samples taken from the posterior distribution are representative and suitable for inference, convergence diagnostics are essential.

### *Gelman-Rubin Diagnostic*

To ensure the accuracy of posterior estimates in Bayesian statistics, it is crucial to assess the convergence of MCMC simulations. The potential scale reduction factor (PSRF), also referred to as the Gelman-Rubin diagnostic, is a popular quantitative technique for evaluating convergence. The volatility inside each chain is compared to the variance across multiple chains. Ideally, the within-chain and between-chain variances should be comparable if every chain is sampling from the same posterior distribution.
We performed several Markov chains for our investigation, beginning them from various, widely separated starting points. After the simulations were finished, we calculated the Gelman-Rubin statistic for each parameter that was of interest, namely for the rate parameter (λ) and the locations of the changepoints (T).

The Gelman-Rubin diagnostics will be presented for every parameter after burn-in in our current study. If the R< statistic is less than 1.1 for every parameter, we will consider the chains to have converged. We will utilise the samples to draw conclusions about the coal mining disaster dataset only after confirming convergence.

### *Traceplots*

When evaluating the convergence of Markov Chain Monte Carlo (MCMC) simulations, traceplots are an essential tool. Plotting the sampled values of a parameter against the iteration number, they are graphical representations. We may visually check the behaviour of the MCMC chains and determine whether they have attained a stationary distribution, which is a sign of convergence, by looking at traceplots.
Each traceplot provides a detailed view of the path taken by the chain through the parameter space. The trace for a well-mixed chain should display a "hairy caterpillar" appearance, indicating that the chain is randomly exploring the distribution of the parameter with no apparent trends or cycles. Key aspects to observe in a traceplot include:

- **Stationarity:** A stationary chain will hover around a certain value with a consistent variance, showing no systematic increase or decrease over time.
- **Mixing:** Good mixing is indicated by a trace that frequently crosses the median value and fills the space of the plot uniformly, without getting "stuck" in any region.
- **Periodicity:** There should be no cyclic or repeating patterns, which would suggest the chain is not exploring the space efficiently.

### *Effecticve Sample Size (ESS)*

The Effective Sample Size (ESS) is a crucial diagnostic metric in Markov Chain Monte Carlo (MCMC) simulations. It measures the quantity of effectively independent samples in a correlated sequence that the MCMC process creates. To assess the accuracy of estimated parameters and ensure the validity of statistical conclusions, ESS is essential. The ESS is computed using the autocorrelation function (ACF) of the MCMC samples for a given parameter. It is defined as:

$$ESS: \frac{N}{1 + 2\sum_{K=1}^{\propto} \rho_k}$$

where N is the total number of MCMC samples and $\rho_k$ is the autocorrelation at lag k.

The samples have little autocorrelation and are almost independent when the ESS value is high and close to the actual sample size. This implies that the chain is investigating the parameter space effectively. A low ESS value, on the other hand, indicates excessive autocorrelation and suggests that fewer truly independent samples were produced, which could compromise the validity of the statistical conclusions.

## Predictive Analysis

In the context of Bayesian models, predictive analysis usually entails projecting future observations using posterior predictive distributions. This method is based on the idea that a distribution that accounts for the uncertainty in the model parameters as well as the intrinsic randomness of the process under study should be used to forecast future data. The following steps will be taken when doing the predictive analysis:

1. **Posterior Estimation:** Utilizing the data already observed, we estimate the posterior distribution of the parameters of our Bayesian model. This involves the implementation of Markov Chain Monte Carlo (MCMC) methods, to generate samples from the posterior distribution.
2. **Predictive Distribution Formation:** Once the posterior distribution is obtained, the next step is to construct the posterior predictive distribution. This is the distribution of a new data point, given the observed data we have. Mathematically, it is obtained by integrating the likelihood of the new data over the entire posterior distribution of the parameters.
3. **Simulation of New Data Points:** With the posterior predictive distribution, we can simulate new data points that are consistent with the patterns and structures identified in the observed data. These simulations will help assess the model's performance and its ability to capture the essential characteristics of the process.

## Model Comparison

A computational method for estimating the marginal likelihood of a statistical model inside a Bayesian framework is called Chib's method, or Chib's Estimation Technique. The marginal likelihood of each model makes it possible to determine which model best fits the observed data, making this method very helpful when comparing several models. Using the Metropolis-Hastings approach, samples are taken from the posterior distribution, and the marginal likelihood is then estimated using those samples.

Chib's method is useful in changepoint detection models because it makes it possible to compare different models that suggest different numbers and locations of changepoints. The goal of changepoint detection models is to identify points in time where the statistical properties of a sequence of observations change. It enables researchers to measure the evidence supporting each model while taking the model's fit to the data and complexity into account.

The formula for Chib's method involves the calculation of the marginal likelihood by integrating the product of the likelihood function and the prior distribution over the entire parameter space. Mathematically, it can be expressed as:

$$p(y) = \int p(y|\theta)p(\theta)d\theta$$

where:

- $p(y)$ is the marginal likelihood of the data.
- $p(y|\theta)$ is the likelihood of the data given the parameters $\theta$.
- $p(\theta)$ is the prior distribution of the parameters.
- $\theta$ represents the vector of the parameters in the model.

Chib's method makes it easier to compare changepoint models in a rigorous and quantitative manner by offering a methodical methodology to estimate the marginal likelihood. It can assist in identifying the most likely quantity and location of changepoints in the historical data on coal mining disasters, which is very helpful for this project as it enhances comprehension of the underlying mechanism and any changes in the frequency of disasters over time.

# No Changepoint Model

To comprehend the fundamental mechanism behind coal mining accidents across time, we investigate a model that asserts a constant event rate—that is, one without changepoints. This strategy is based on the assumption that the frequency of disasters is constant, without of any notable fluctuations that could indicate external interventions or structural modifications within the industry.

## Model Implementation

For the Metropolis-Hastings algorithm, we initialised four different chains in order to guarantee robustness and lessen the impact of starting values. Randomly selected from the previous distribution, each chain started at a different place to offer a variety of viewpoints on the parameter space.

With a normal proposal distribution cantered at the present value of λ and a standard deviation of 0.1, the Metropolis-Hastings algorithm was described in detail. In order to efficiently explore the parameter space and maintain a reasonable acceptance rate for proposed samples, this proposal distribution was selected.

Each of the four chains was run for 10,000 iterations to ensure a thorough exploration of the posterior landscape. We assessed the need for a burn-in period by visual inspection of the traceplots, ensuring that the chains had reached a stationary distribution before collecting samples for analysis.

Posterior Distribution of the model:

$$Gamma(\alpha + \sum_{i=1}^{112} y_i,\ \beta + 112), where\ \alpha = 2\ and\ \beta = 1$$

## Prior and Posterior Density Plots

The previous understanding of the disaster rate parameter was represented graphically in Figure 2, where a red line represented the density of the gamma distribution over a range of values. After integrating the data, the posterior distribution was calculated and plotted, showing a revised belief about based on the actual disaster counts as represented by the blue line in Figure 3. The posterior plot shows how the data were analysed, moving from a prior distribution based on historical knowledge to a posterior distribution based on empirical knowledge.

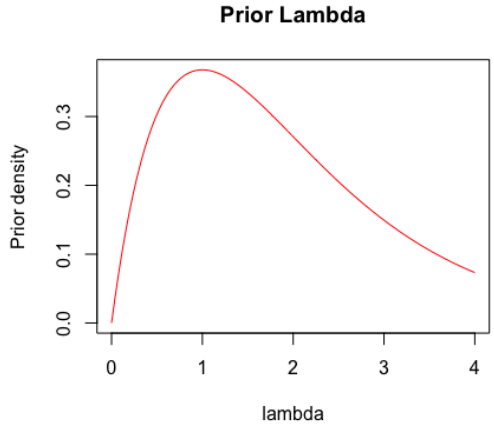


**Figure 2: Prior Lambda Density**
A plot showing the gamma prior distribution for the rate parameter $\lambda$ ,indicating initial beliefs about the disaster

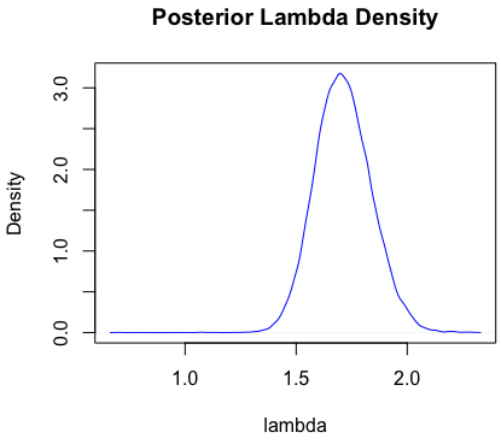


**Figure 3: Posterior Lambda Density**
The posterior density plot of $\lambda$, illustrating the updatedd distribution after data integration, showing the Bayesian

### *Traceplot*

Traceplots were created for each of the four chains so that the convergence could be visually inspected. The sampled values of the rate parameter λ throughout the iterations are shown in these plots. A strong sign of convergence can be seen in Figure 5 where the chains' paths overlap and mix, indicating that they are exploring the same areas of the parameter space.

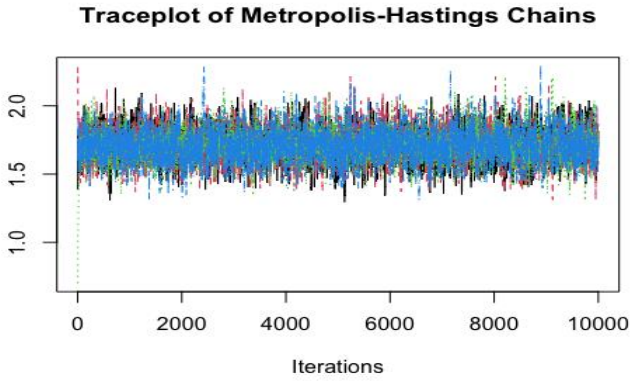


**Figure 4: Traceplots of MCMC Chains for the Rate Parameter λ**
The figure illustrates the traceplots for four Metropolis-Hastings chains used to sample the posterior distribution of the rate parameter λ. Each chain is represented by a different colour, and their intermingling suggests that the chains have reached a stationary distribution, a key criterion for convergence.

### *Gelman-Rubin Diagnostic*

We used the Gelman-Rubin diagnostic in addition to visual inspection to statistically evaluate convergence. By using this method, the variance within chains is compared to the variance between chains. Convergence is implied when the Gelman-Rubin statistic is near 1. This indicates that intra- and inter-chain variance are comparable. Gelman-Rubin statistics at or very close to 1 were obtained for all parameters by our analysis, which increased our confidence in the chains' convergence.

**Table 1: Gelman-Rubin Diagnostic Results for No changepoint**

| **Potential scale reduction factors:** | |
|---|---|
| **Point est.** | Upper C.I. |
| **1** | 1.01 |

**Table 1:** Gelman-Rubin diagnostic results for assessing the convergence of MCMC chains. The Potential Scale Reduction Factor (PSRF), also known as the Gelman-Rubin statistic $\hat{R}$, for the rate parameter $\lambda$ is shown. A point estimate of 1.00 and an upper confidence interval of 1.01 suggest that the chains have converged to the target posterior distribution.

# Single Changepoint Model

## Model Implementation

We use a single changepoint model to try and find structural changes in our time series data. This strong statistical method is intended to identify a particular time point, T, at which a significant change in the underlying event rate λ takes place. By doing this, we can gain a deeper comprehension of the underlying mechanisms and dynamics influencing the events that are observed. The model is characterized by two distinct Poisson rates:

- $\lambda_1$: The rate before the changepoint.
- $\lambda_2$: The rate after the changepoint.

To initialize the model, we set $\lambda_1$and $\lambda_2$ equal to the overall mean of observed event counts. This choice serves as a neutral starting point for the MCMC algorithm. For the changepoint T, we select the time corresponding to the peak of observed counts as our initial guess. This heuristic approach assumes that a significant change in the event rate is likely to cause a noticeable spike in event frequency.

We sample from the posterior distributions of our parameters using the Metropolis-Hastings algorithm. The algorithm's effectiveness is evaluated using acceptance ratios that are derived from several MCMC chains. The acceptance ratios of 0.416, 0.408, 0.403, and 0.405 in our analysis point to a well-functioning proposal mechanism. These ratios exhibit a high likelihood of convergence to the true posterior distribution and show an effective exploration of the parameter space, all while falling comfortably within the desired range.

### *Posterior Distributions*

The posterior distributions for $\lambda_1$ and $\lambda_2$ are modelled as Gamma distributions with shape and rate parameters updated based on the observed data. The formulas are as follows:

Posterior distribution of $\lambda_1$:

$$Gamma\left(\alpha + \sum_{i=1}^{T-1} y_i\,, \beta + (T-1)\right), where\ \alpha = 2\ and\ \beta = 1$$

Posterior distribution of $\lambda_2$:

$$Gamma\left(\alpha + \sum_{i=T}^{112} y_i\,, \beta + (112-T)\right), where\ \alpha = 2\ and\ \beta = 1$$

Posterior of T:

$$\propto ((T-1)+1)^{-2-\sum_{I=1}^{T_1} y_i} \Gamma\left(2 + \sum_{i=1}^{T_1-1} y_i\right) + \left((112-(T-1)\right)+1)^{-2-\sum_{I=T}^{112} y_i}\, \Gamma\left(2 + \sum_{i=1}^{T_1-1} y_i\right)$$

## Prior and Posterior Density Plots

**Posterior Distribution of $\lambda_1$:** The unimodal and peaked nature of $\lambda_1$'s posterior density, as depicted in Figure 5, is indicative of data providing substantial insight, yielding a precise estimate of the pre-changepoint event rate. This level of estimation certainty is indicative of the robustness of our model in capturing the dynamics prior to the changepoint.

**Joint Density of $\lambda_1$ and $\lambda_2$:** The concentration of sample density in the joint plot of $\lambda_1$ and $\lambda_2$ suggests a correlation between the event rates across the changepoint. A strong correlation, inferred from tightly packed contour lines, highlights a systematic shift in event rates, reaffirming the significance of the changepoint in our analysis.

**Joint Density of $\lambda_1$ and T:** The joint plot between $\lambda_1$ and the changepoint T can reveal the consistency of the changepoint's timing relative to the pre-changepoint event rate. Strong high-density region clustering highlights the validity of the changepoint identification and its impact on $\lambda_1$.

**Posterior Distribution of $\lambda_2$**: The plot for $\lambda_2$ illustrates the post-changepoint event rate with a high degree of precision if a sharp peak is observed. Accurate estimation of $\lambda_2$ is critical, as it potentially reflects the impact of interventions occurring at the changepoint.

**Contour Plot of $\lambda_2$ and T**: The contour plot that includes$\lambda_2$ and T examines how they interact, with different patterns suggesting different post-changepoint behaviours. These patterns could draw attention to factors affecting the event rate that are not visible until after the changepoint, strengthening the explanatory capacity of our model.

**Histogram of T:** Changepoint estimates from all MCMC samples are combined into a histogram for T, where a prominent peak indicates a clearly defined changepoint. For the purpose of

identifying changes in the process being observed, changepoint estimation's clarity is essential. A histogram with several peaks or a wide distribution, on the other hand, would indicate the need for more information or model improvement.

**Parameter Distribution and Correlation in the Single Changepoint Model**

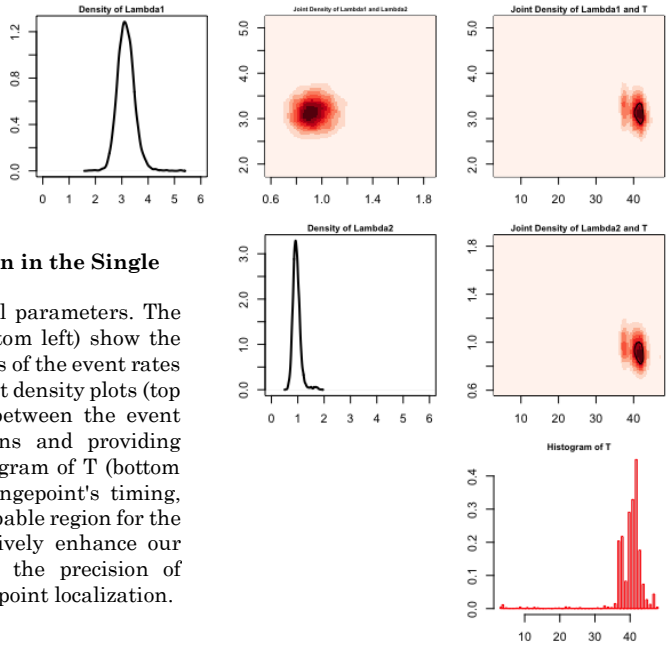


**Figure 5: Parameter Distribution and Correlation in the Single Changepoint Model**
Multifaceted analysis of the single changepoint model parameters. The univariate density plots for λ1 (top left) and λ2 (bottom left) show the posterior distributions, indicating the refined estimates of the event rates before and after the changepoint, respectively. The joint density plots (top and middle right) illustrate the interdependencies between the event rates and the changepoint, highlighting correlations and providing insight into their conditional relationships. The histogram of T (bottom right) presents the posterior distribution of the changepoint's timing, with its kernel density overlay indicating the most probable region for the changepoint occurrence. These visualizations collectively enhance our understanding of the model's dynamics, informing the precision of parameter estimates and the confidence in the changepoint localization.

## Convergence Diagnostics

### *Gelman-Rubin Diagnostic*

**Table 2: Gelman-Rubin Diagnostic Results for Single changepoint**

| **Potential scale reduction factors:** | |
|---|---|
| **Point est.** | Upper C.I. |
| **1** | 1.01 |
| **1** | 1.00 |
| **1** | 1.01 |
| **Multivariate psrf.** | |
| **1** | |

**Table 2: Gelman-Rubin Diagnostic Results**
This table presents the potential scale reduction factors (PSRF) for the parameters of the single changepoint model. Values close to 1.00 for each parameter—Lambda1, Lambda2, and T—indicate that convergence has likely been reached. The upper confidence interval remains close to 1, further supporting the adequacy of the convergence. The multivariate PSRF value of 1 demonstrates overall model convergence, suggesting that the parameter estimates are consistent across multiple MCMC chains.

All parameters, including the changepoint T, have PSRF values close to 1, indicating that our chains have stabilised and are yielding reliable estimates in different runs. This is a strong signal that our posterior distributions are robust, allowing us to proceed with high confidence in our interpretation of the MCMC results. The robustness of our single changepoint model is reinforced by the chains' convergence, as indicated by the Gelman-Rubin diagnostic.

### *Traceplots*

The trace plots (Figure 6) for the event rate parameters before $\lambda_1$ and after $\lambda_2$ the changepoint illustrate the path of the MCMC sampler through the parameter spaceThe fact that different chains' traces overlap indicates that they are exploring similar regions and mixing well, both of which point to convergence towards a stationary distribution. In the absence of any discernible trend or periodic patterns, non-convergence or problems with the sampling algorithm would be suspected.

Across all chains, the trace plot for the changepoint parameter T shows a consistent sampling pattern that stays within the same range of values. The absence of drift in the T traces suggests that the chains have become acclimated to the posterior distribution and that the timing of the changepoint is being estimated consistently over several runs.

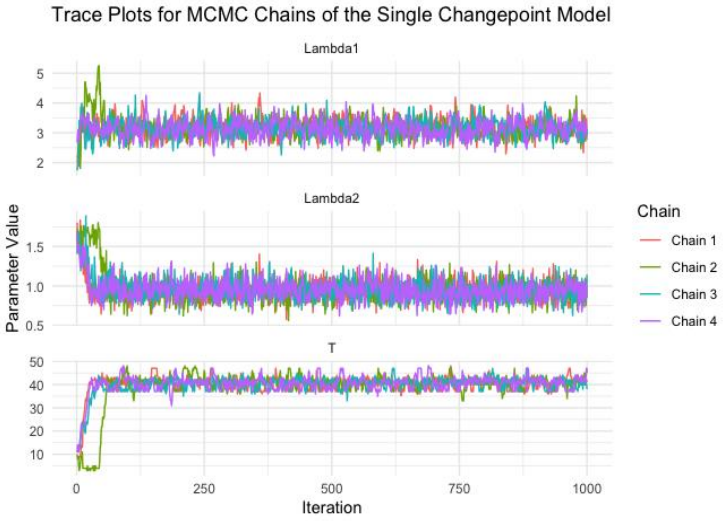


**Figure 6:** This series of trace plots displays the sampled values for parameters $\lambda_1$, $\lambda_2$, and T across 1,000 iterations of the MCMC simulation, stratified by chain. The plots reveal the convergence behaviour of each parameter, with the chains for $\lambda_1$ and $\lambda_2$ demonstrating consistent mixing and stability, indicative of good convergence. The trace for T shows that the chains remain within a specific range, suggesting a consensus on the changepoint's location. The lack of visible trends or cycles within the traces supports the conclusion that the chains have reached equilibrium, validating the use of these samples for posterior inference.

## Effective Sample Size

**Table 3: ESS Two Changepoint Model**

| $\lambda_1$ | $\lambda_2$ | T |
|---|---|---|
| 1098.14 | 463.13 | 182.03 |

**Table 3:** ESS for each parameter of the single changepoint model, quantifying the number of independent-like samples from the correlated MCMC chains. An ESS of 1098.14 for λ1 suggests a high degree of sample independence, while the lower ESS for λ2 (463.13) and T (182.03) indicates more pronounced autocorrelation, suggesting the need for improved sampling strategies for these parameters.

The Effective Sample Size (ESS) for the parameters in our model is shown in Table 3. In our one-changepoint model, $\lambda_1$ has an ESS greater than 1000, indicating a sample set that is highly independent, supporting the validity of our posterior inferences. On the other hand, $\lambda_2$ and T exhibit lower ESS values (463.13 and 182.03, respectively), indicating a higher degree of autocorrelation and a requirement for improved sampling efficiency.

For simple analyses, the ESS for T and $\lambda_2$ might be sufficient, but for stable estimates, best practices suggest a higher threshold. Improvements to the MCMC procedure for these parameters may help to further improve the accuracy and dependability of the model's conclusions.

# Two Changepoint Model

## Model Implementation

The Metropolis-Hastings algorithm is initialized by setting the parameters, $\lambda_1$, $\lambda_2$, and $\lambda_3$, to the mean of the observed data. Changepoints, $T_1$ and $T_2$, are strategically placed at one-third and two-thirds of the timeline, respectively, providing an equitable initial exploration of the parameter space.

To facilitate efficient sampling, the algorithm updates the parameters using a symmetric normal proposal distribution, with a standard deviation set to 0.1 for λ parameters and 2 for T parameters. The requirement that $T_1$ come before $T_2$is a crucial restriction; any proposal that doesn't meet this requirement is automatically turned down. The likelihood ratio test, which considers the previous distributions, determines whether to accept new proposals.

10,000 iterations of the model are performed, with a thinning factor of 5 added to reduce autocorrelation. This yields 2,000 samples for every chain. Four independent chains are run to verify convergence, and the following acceptance ratios are obtained:

1. Chain 1 Acceptance Ratio: 31.87%
- Chain 2 Acceptance Ratio: 33.24%
- Chain 3 Acceptance Ratio: 33.25%
- Chain 4 Acceptance Ratio: 33.51%

These ratios align with the optimal range of 20-50%, indicating an effective balance between exploration breadth and acceptance rate.

### *Posterior Distributions*

The posterior distributions for the λ parameters are modelled as Gamma distributions, updated based on the observed data segmented by the changepoints:
Posterior distribution of $\lambda_1$:

$$Gamma\left(\alpha + \sum_{i=1}^{T_1-1} y_i\,, \beta + (T_1 - 1)\right), where\ \alpha = 2\ and\ \beta = 1$$

Posterior distribution of $\lambda_2$:

$$Gamma\left(\alpha + \sum_{i=T_1}^{T_2-1} y_i\,, \beta + (T_2 - T_1)\right), where\ \alpha = 2\ and\ \beta = 1$$

Posterior distribution of $\lambda_3$:

$$Gamma\left(\alpha + \sum_{i=T_2}^{112} y_i\,, \beta + (112 - (-T_2 - 1))\right), where\ \alpha = 2\ and\ \beta = 1$$

Posterior of T's:

$$\propto ((T_1 - 1) + 1)^{-2-\Sigma_{i=1}^{T_1-1} y_i}\Gamma\left(2 + \sum_{i=1}^{T_1-1} y_i\right) + ((T_2 - 1 - T_1) + 1)^{-2-\Sigma_{i=T_1}^{T_2-1} y_i}\Gamma\left(2 + \sum_{i=T_1}^{T_2-1} y_i\right)$$

$$+\left(\left(112 - (T_2 - 1)\right) + 1\right)^{-2-\Sigma_{I=T_2}^{112} y_i}\,\Gamma\left(2 + \sum_{i=T_2}^{112} y_i\right)$$

The posterior of the changepoints, $T_1$and $T_2$, is proportional to the product of the likelihoods of observing the data within each segment defined by the changepoints, penalizing configurations that fail to explain the observed data well.

## Prior and Posterior Density Plots

The univariate density plots at the top for the parameters $\lambda_1$, $\lambda_2$, and $\lambda_3$ are shown in Figure 7. These densities, which have discrete peaks, show accurate posterior estimates for event rates over the time series that the changepoints have divided. These peaks' sharpness confirms that the data had a significant impact on improving our priors and producing accurate posterior estimates.
Joint density plots for pairs $\lambda_1$ and $\lambda_2$, and $\lambda_1$, $\lambda_3$, and $\lambda_2$ and $\lambda_3$ are shown in the middle of Figure 7. The plots show that event rates are positively correlated and that lower rates, particularly near changepoints, are more in line with the data that has been observed.
At the bottom, the 3D plot elucidates the model's selection mechanism for changepoints $T_1$ and $T_2$. Peaks in this plot represent the most credible years for changepoints as discerned from the data. The varied heights of these peaks reflect the acceptance rates during sampling, with taller peaks

indicating a higher acceptance frequency. This sensitivity to changepoint placement is typical and desired in changepoint analysis.

**Figure 7: Prior and Posterior Density Visualization for Two Changepoint Model**

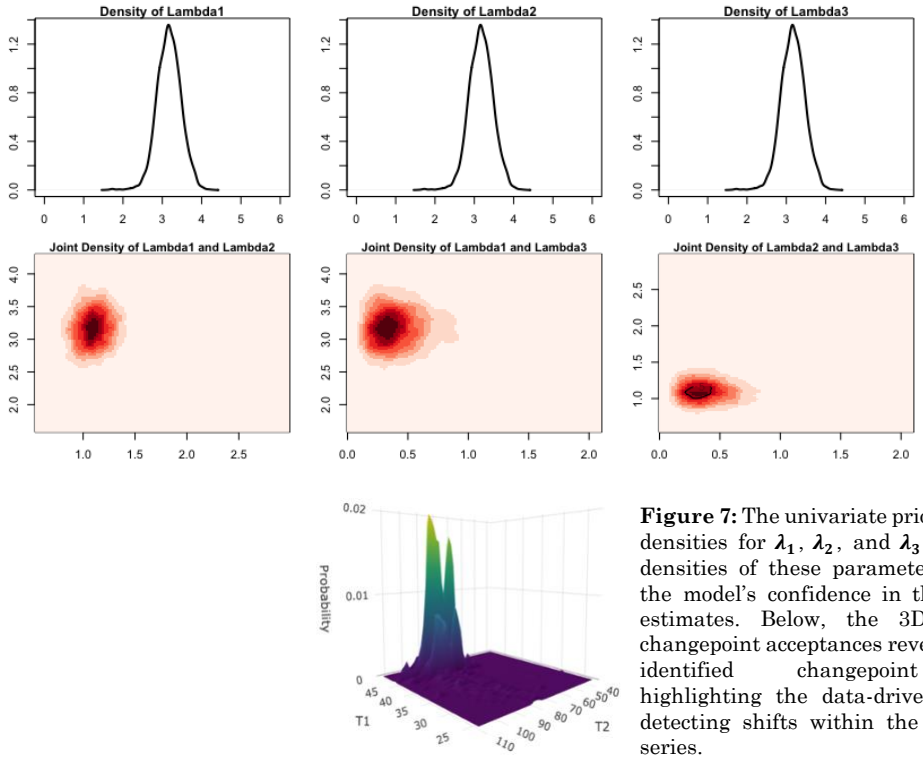


**Figure 7:** The univariate prior and posterior densities for $\boldsymbol{\lambda_1}$, $\boldsymbol{\lambda_2}$, and $\boldsymbol{\lambda_3}$, and the joint densities of these parameters, showcasing the model's confidence in the event rates' estimates. Below, the 3D plot of the changepoint acceptances reveals the model's identified changepoint positions, highlighting the data-driven prevision in detecting shifts within the observed time series.

## Convergence Diagnostics

### *Gelman- Rubin Diagnostic*

The individual parameter PSRF values for our two changepoint model span 1.03 to 1.24, while the upper confidence intervals (Upper C.I.) cover 1.08 to 1.69. These numbers suggest that some parameters may not be convergent, especially in cases where the upper confidence interval is noticeably larger than 1.0. Simultaneous consideration of all parameters yields a slightly non-convergent multivariate PSRF of 1.06.

The majority of the parameters, according to the diagnostic results, are on the verge of convergence; however, more iterations might be required to guarantee that all parameters have fully converged, especially for the parameter with a PSRF value of 1.24. When the PSRF values are higher than 1.0, even slightly, care should be taken. They imply that the chain variance is slightly larger than the within-chain variance, suggesting that continued sampling could potentially alter the posterior distributions of the parameters.

**Table 4: Gelman-Rubin Diagnostic Results for Two changepoints**

| **Potential scale reduction factors:** | |
|---|---|
| **Point est.** | Upper C.I. |
| **1.03** | 1.08 |
| **1.05** | 1.07 |
| **1.06** | 1.12 |
| **1.00** | 1.01 |
| **1.24** | 1.69 |
| **Multivariate psrf.** | |
| **1.06** | |

**Table 4 :**Table of Gelman-Rubin Diagnostic Values for the Two Changepoint Model indicating the range of PSRF values for individual parameters and their upper confidence intervals. A PSRF value close to 1 suggests convergence, while values substantially greater than 1 indicate non-convergence. The table reflects that most parameters are approaching convergence, but a few exhibit PSRF values that warrant further iterations to achieve convergence across all parameters, with special attention needed for the parameter with the highest PSRF value.

Further MCMC runs with increased iterations, potentially with adjusted tuning parameters for the proposal distribution, may enhance convergence. Additionally, examining the trace plots for individual chains can provide further insight into the mixing and stationarity of the chains.

## *Traceplots*

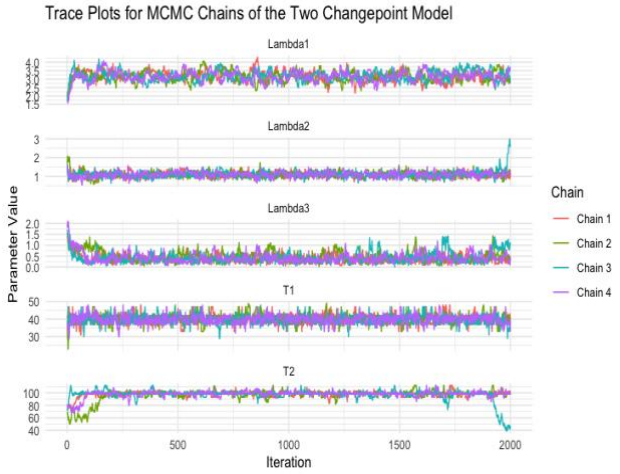


**Figure 8:** Traceplots depicting the parameter estimates for the Two Changepoint Model across 2,000 iterations. Each plot shows the trajectories of four independent MCMC chains for the rate parameters ( $\boldsymbol{\lambda_1}$ , $\boldsymbol{\lambda_2}$ , $\boldsymbol{\lambda_3}$ ) and changepoint parameters ($\boldsymbol{T_1}$, $\boldsymbol{T_2}$). Consistent overlapping of the chains for the Lambda parameters suggests convergence, while the divergence in one of the chains for T2 in later iterations may indicate potential areas for further exploration to ensure robust model convergence.

The traceplots for the rate parameters $\lambda_1$, $\lambda_2$, and $\lambda_3$ show stable, well-mixed chains, indicating the achievement of a stationary distribution. Reliability in convergence is indicated by the consistency of overlapping traces among the four chains, and the posterior distribution appears well-defined.
$T_1$ shows some variability in convergence for the changepoints, while $T_2$ shows a shift in one chain as the iteration comes to an end. This may indicate possible problems with convergence or a new discovery in the parameter space. If exclusive to a single chain, it could point to an undiscovered region with a high posterior density, which would require further investigation to ascertain the nature of this shift.

In conclusion, the MCMC algorithm successfully converges for λ parameters; however, additional research is necessary to verify the significance of the shift that was noticed in the $T_2$ traceplot.

## *Effective Sample Size*

**Table 5: ESS – Two Changepoint Model**

| $\boldsymbol{\lambda_1}$ | $\lambda_2$ | $\lambda_3$ | $T_1$ | $T_2$ |
|---|---|---|---|---|
| 350.37 | 1015.74 | 399.50 | 1283.67 | 149.69 |

**Table 5** : Table displaying the Effective Sample Size (ESS) for parameters of the Two Changepoint Model. ESS values signify the quality of MCMC sampling, with higher numbers indicating better parameter estimation. The table shows a range of ESS values across parameters, with $\boldsymbol{T_2}$ requiring attention due to its notably low ESS, suggesting the need for improved sampling efficiency or additional iterations to ensure robust model convergence and reliable parameter estimates.

The ESS values of our model's parameters vary, indicating different levels of sampling effectiveness. Notably high ESS values (1015.74 and 1283.67 for $\boldsymbol{\lambda_2}$ and $\boldsymbol{T_1}$, respectively) indicate robust parameter estimation. Nevertheless, the ESS values of $\boldsymbol{\lambda_1}$, $\boldsymbol{\lambda_3}$, and $\boldsymbol{T_2}$ are lower; $\boldsymbol{T_2}$'s value is only 149.69, indicating significant autocorrelation and possible difficulties with sampling, especially in precisely identifying the second changepoint.
The accuracy of the model and the dependability of its conclusions may be impacted by these variations in ESS. Because the low ESS for $\boldsymbol{T_2}$ affects the determination of changepoints, addressing it is essential. Techniques like increasing the number of iterations, fine-tuning the tuning, or changing the proposal distribution can enhance ESS and, consequently, the general validity of the model.

# Three Changepoint Model

## Model Implementation

To establish objectivity, the model starts with the λ parameters set to the mean of the data. Changepoints (T parameters) are first distributed throughout the data span into equal quartiles, providing a methodical starting point for changepoint placement.
To ensure complete parameter space coverage, standard deviations for updates to the λ parameters are derived from previous analytics and follow a symmetric normal proposal distribution. The T parameters take into consideration the intrinsic uncertainty in identifying several changepoints by using a wider proposal distribution.
Under the constraint $T_1 < T_2 < T_3$, sequential integrity is required, and proposals that do not comply are rejected. Acceptance of new proposals is determined by a likelihood ratio that includes uniform priors for T and Gamma (2,1) priors for λ.
The algorithm has a burn-in phase for estimate stabilisation and is highly iterated, outperforming the Two Changepoint Model. A thinning factor is applied as needed to reduce autocorrelation and promote sample independence.
Across four independent chains, the algorithm ensures thorough convergence checks, with acceptance ratios optimally set between 20%-50%. These values indicate adept navigation of the parameter landscape:

- Chain 1 Acceptance Ratio: 29.24%
- Chain 2 Acceptance Ratio: 30.39%
- Chain 3 Acceptance Ratio: 31.64%
- Chain 4 Acceptance Ratio: 31.97%

### *Posterior Distributions*

The posterior distributions for λ are modelled as Gamma distributions with parameters α and ß set to 2 and 1, respectively. Each λ parameter is updated based on the data within the intervals demarcated by the T parameters.

Posterior distribution of $\lambda_1$:

$$Gamma\left(\alpha + \sum_{i=1}^{T_1-1} y_i, \beta + (T_1 - 1)\right), where\ \alpha = 2\ and\ \beta = 1$$

Posterior distribution of $\lambda_2$:

$$Gamma\left(\alpha + \sum_{i=T_1}^{T_2-1} y_i, \beta + (T_2 - T_1)\right), where\ \alpha = 2\ and\ \beta = 1$$

Posterior distribution of $\lambda_3$:

$$Gamma\left(\alpha + \sum_{i=T_1}^{T_3-1} y_i, \beta + (T_3 - T_2)\right), where\ \alpha = 2\ and\ \beta = 1$$

Posterior distribution of $\lambda_4$:

$$Gamma\left(\alpha + \sum_{i=T_2}^{112} y_i, \beta + (112 - (-T_3 - 1))\right), where\ \alpha = 2\ and\ \beta = 1$$

The T parameters' posterior is a function of the preceding data points, with a penalizing factor for lesser data support.

Posterior of T's:

$$\propto ((T_1 - 1) + 1)^{-2-\sum_{i=1}^{T_1-1} y_i} \Gamma\left(2 + \sum_{i=1}^{T_1-1} y_i\right) + ((T_2 - 1 - T_1) + 1)^{-2-\sum_{i=T_1}^{T_2-1} y_i} \Gamma\left(2 + \sum_{i=T_1}^{T_2-1} y_i\right)$$

$$+ ((T_3 - 1 - T_2) + 1)^{-2-\sum_{i=T_2}^{T_3-1} y_i} \Gamma\left(2 + \sum_{i=T_2}^{T_3-1} y_i\right)$$

$$+ \left(\left(112 - (T_2 - 1)\right) + 1\right)^{-2-\sum_{i=T_3}^{112} y_i} \Gamma\left(2 + \sum_{i=T_2}^{112} y_i\right)$$

## Convergence Diagnostics

### *Gelman-Rubin Diagnostic*

Table 6's Gelman-Rubin diagnostics indicate that the parameters of the Three Changepoint Model have good convergence. The model parameters show strong signs of convergence with PSRF values between 1.02 and 1.13 and upper confidence limits between 1.05 and 1.23. Values close to 1.0 indicate consistent variance both within and across MCMC chains.

| **Potential scale reduction factors:** | |
|---|---|
| **Point est.** | Upper C.I. |
| **1.02** | 1.05 |
| **1.04** | 1.09 |
| **1.05** | 1.13 |
| **1.01** | 1.02 |
| **1.03** | 1.05 |
| **1.04** | 1.12 |
| **1.13** | 1.23 |
| **Multivariate psrf.** | |
| **1.06** | |

**Table 6:** Table of Gelman-Rubin Diagnostic Results for the Three Changepoint Model, displaying potential scale reduction factors (PSRF) for the λ and T parameters. Values close to 1 suggest that convergence has likely been achieved. The multivariate PSRF, slightly above 1, points to an adequate level of convergence across all parameters, with some individual parameters indicating a need for further analysis.

The model's multivariate PSRF at 1.06 confirms that the chains are mixing adequately, pointing to a collectively reliable estimation of parameters. One parameter, with a slightly higher PSRF and Upper C.I., may require further iterations to ensure thorough convergence.

### *Traceplots*

The Three Changepoint Model's traceplots display a range of outcomes. $\boldsymbol{\lambda_1}$ and $\boldsymbol{\lambda_2}$ show convergence, but $\boldsymbol{\lambda_3}$ and $\boldsymbol{\lambda_4}$ show variability, which may indicate incomplete convergence. The traces of the changepoint parameters point to identifiability issues, which are critical for precise changepoint location and model predictions.
These results underline the necessity of further iterations or possible improvements to the MCMC algorithm to enhance parameter estimation and guarantee the predictive integrity of the model.

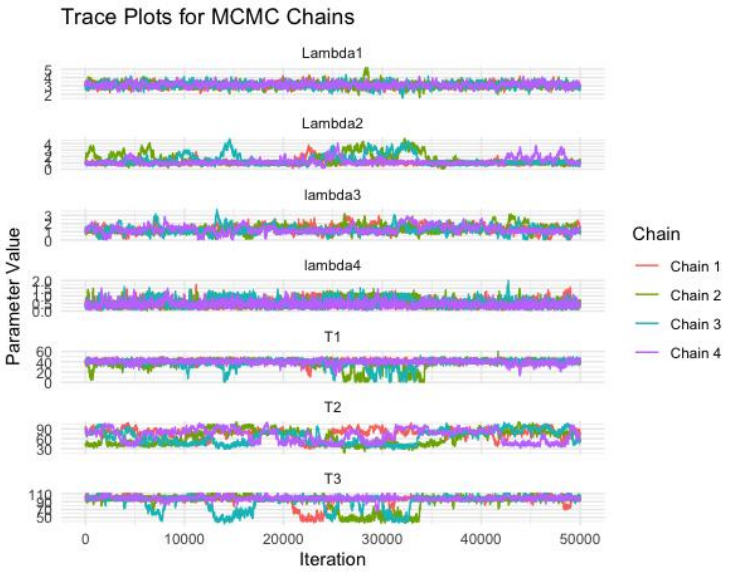


**Figure 9:** Traceplots for MCMC Chains of the Three Changepoint Model over 50,000 iterations, showing the behavior of four chains for each parameter: $\boldsymbol{\lambda_1}$, $\boldsymbol{\lambda_2}$, $\boldsymbol{\lambda_3}$, $\boldsymbol{\lambda_4}$, $\boldsymbol{T_1}$, $\boldsymbol{T_2}$, and $\boldsymbol{T_3}$. The plots reveal varying degrees of stability and convergence across the parameters, with some exhibiting consistent convergence while others, particularly the T parameters, display divergence and variability indicative of complex dynamics in the parameter space.

### *Effective Sample Size*

**Table 7: ESS – Three Changepoint Model**

| $\boldsymbol{\lambda_1}$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $T_1$ | $T_2$ | $T_3$ |
|---|---|---|---|---|---|---|
| 1152.77 | 135.83 | 167.25 | 700.85 | 304.44 | 65.99 | 408.60 |

**Table 7:** Table of Effective Sample Size (ESS) for the Three Changepoint Model, detailing the number of independent-like samples for each parameter. The ESS values highlight the varying levels of sampling efficiency across parameters, with some (like $\lambda_1$) showing high independence and others (particularly $T_2$) indicating potential issues due to lower ESS and the need for improved sampling strategies.

The difficulties in estimating multiple changepoints and rate parameters within the Three Changepoint Model are highlighted by the variation in ESS across different parameters. In particular, the low ESS values for $T_2$, $\lambda_2$, and $\lambda_3$ are alarming because they imply that these parameters are not being estimated with the same level of reliability as others. The predictive accuracy and overall performance of the model may be strongly impacted by these parameters. As a result, it might be essential to think about methods for enhancing sampling efficiency, like changing the proposal distribution, doing more iterations overall, or looking into different MCMC algorithms.

In conclusion, the Three Changepoint Model exhibits a good degree of sampling efficiency for certain parameters, but other parameters obviously need to be addressed to guarantee the model's robustness and the accuracy of its conclusions. The low ESS for certain parameters, particularly for $T_2$, points to the need for further diagnostics and potentially additional MCMC runs to achieve more precise estimations.

# Four Changepoint Model

## Model Implementation

In the four changepoint model, we aim to identify shifts in event rates over time, hypothesizing that these changes occur at four specific points. We model the event rates using a piecewise constant function, where each segment's rate is distinct and is parameterized by λ, and the shifts are defined by the changepoints T. The posterior distributions for those parameters are derived using Bayesian inference.

1) Distributions of Event Rates (λ) in the posterior: A Gamma distribution governs each rate parameter λ, and it is updated in accordance with the number of events (y) observed within the intervals marked by the changepoints (T). The following is a description of the event rates' gamma distributions:

   - For the first segment, before the first changepoint $T_1$, the rate $\lambda_1$ is updated using all events up to $T_1$-1.
   - For subsequent segments, each λ is updated using the event counts between its bounding changepoints.
   - The shape parameter α is set to 2, and the rate parameter $\beta$ is set to 1 for all the λ's, ensuring consistency in the prior beliefs across all segments.

2) Posterior Distribution for Changepoints (T): The posterior for the changepoints T is a complex function dependent on the event rates λ and the observed events y. It's not a standard distribution but rather a likelihood function multiplied by a prior, if available. The Metropolis-Hastings algorithm samples from this posterior, accepting new changepoint proposals that increase the posterior probability.

When we apply the four changepoint model's implementation section with these posterior distributions, we obtain:

The model assumes that the posterior distributions of the event rates $\lambda_1$, $\lambda_2$, $\lambda_3$, $\lambda_4$, and $\lambda_5$ follow Gamma distributions, with parameters updated according to the number of events within each interval. The rates are estimated from the data bounded by the changepoints. The first rate, $\lambda_1$, is updated from the beginning of the data to immediately prior to the first changepoint, $T_1$, and the other rates are updated in a similar manner.

To represent the moments at which the event rate changes, the changepoints $T_1$, $T_2$, $T_3$, and $T_4$ are modelled. Their posterior is based on the probability of the observed data up to that point, with any given prior for the changepoints and a Gamma-shaped prior for the event rates serving as guidance.

Iteratively fine-tuning the estimates for both the event rates λ and the changepoints T to fit the observed data with a high degree of accuracy ensures the convergence and reliability of the model.

The histograms depicted in Figure 10 were generated to visualize the marginal posterior distributions of the changepoints $T_1$, $T_2$, $T_3$ and $T_4$in a four-changepoint model. These distributions were obtained from 60,000 MCMC iterations, and the prominent spikes in the histograms represent the years where changepoints are most probable according to the posterior distribution.

These spikes imply that there is substantial evidence in the data for changes in the underlying process during these periods. The years denoted by the peaks suggest that there was probably a shift in the rates of the events or phenomena under study during these times. Such spikes might point to years when major technological advancements or changes in safety regulations occurred, for example, if the model is tracking the rate of industrial accidents.

We have great confidence in the estimated location of the changepoints for our model because these peaks are clearly defined. It shows that the model has a high degree of precision in identifying structural breaks in the data. These findings are critical for understanding the temporal dynamics in the data and can be instrumental in decision-making processes that rely on identifying periods of significant change.

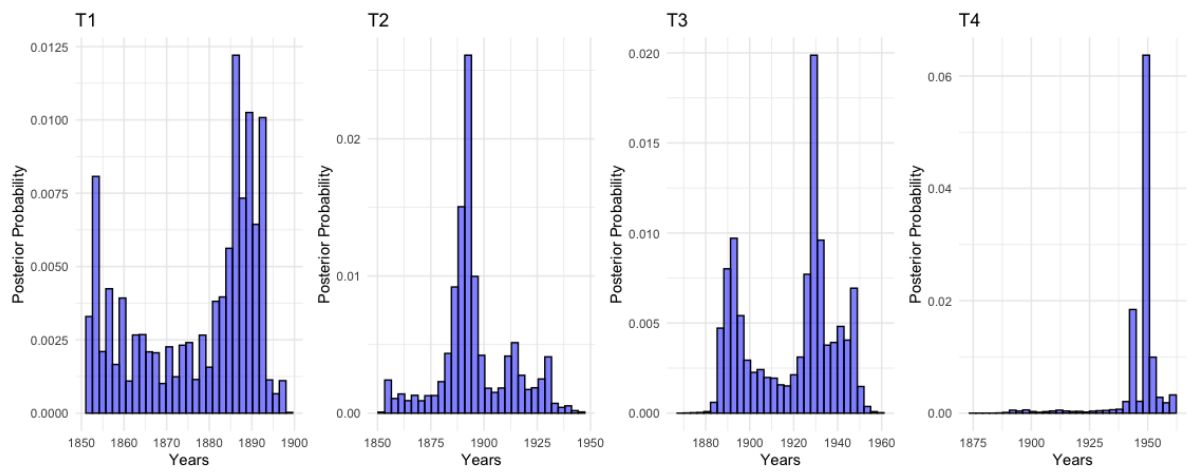


**Figure 10:** Based on the 60,000 iterations from each of the marginal (conditional posterior distribution of $T_1$, $T_2$, $T_3$ and $T_4$. Strong spike in the histogram indicates where the changepoints are likely to be located. Therefore $T_1$ seems to be most likely located between 1885 to 1894; $T_2$ between 1888 to 1897; $T_3$ between 1930 to 1933; and $T_4$ between 1943 to 1950.

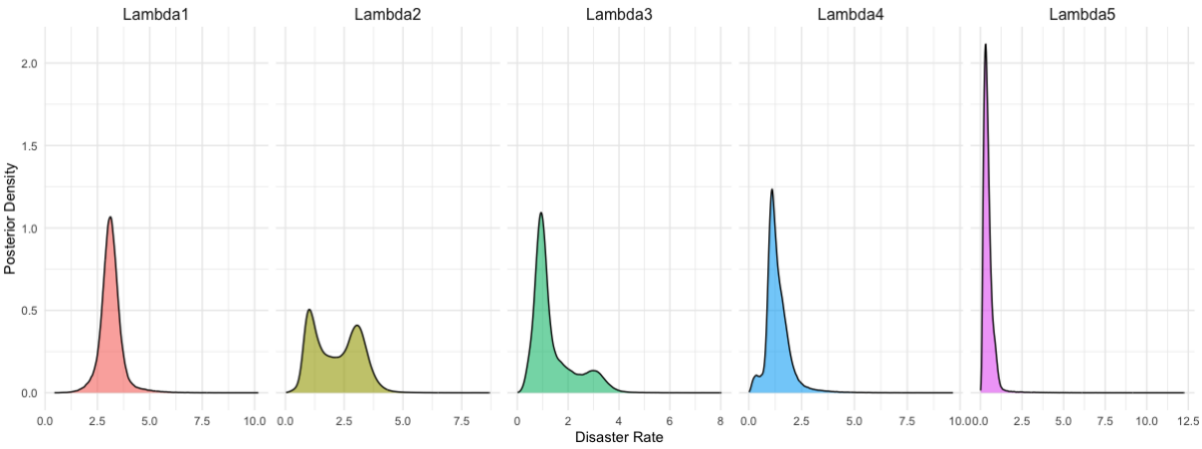


**Figure 11:** Panels show density plots of 60,000 samples from the conditional posterior distribution of $\lambda_1$, $\lambda_2$, $\lambda_3$, $\lambda_4$ and $\lambda_5$ respectively. Strong spike in density plots of $\lambda_1$, $\lambda_3$, $\lambda_4$ and $\lambda_5$ indicates what the rate of yearly number of coal mining disasters is likely to be for the corresponding segment. Therefore, the value of $\lambda_1$ is likely to be between 2.5 and 3.4; $\lambda_2$ between 0.5 and 5; $\lambda_3$ between 0.5 and 1.7; $\lambda_4$ between 0.6 and 1.4; $\lambda_5$ is between 0 and 1.25.

Figure 11 presents the density plots for the conditional posterior distributions of the event rate parameters $\lambda_1$, $\lambda_2$, $\lambda_3$, $\lambda_4$ and $\lambda_5$ in a model designed to estimate the annual number of coal mining disasters. These plots were derived from 60,000 samples, and the distinct spikes observed in the densities of $\lambda_1$, $\lambda_3$, $\lambda_4$ and $\lambda_5$ provide a clear indication of the most probable event rates for their respective time segments. The values of $\lambda_1$ are most likely to fall between 2.5 and 3.4, $\lambda_3$ between 0.5 and 1.7, $\lambda_4$ between 0.6 and 1.4, and $\lambda_5$ between 0 and 1.25.

Interestingly, there is no discernible peak in the density plot for $\lambda_2$, suggesting a greater degree of uncertainty or variability in the event rate for this portion of the timeline. The lack of a clear peak may indicate a phase of instability or transition in the process being modelled, which could be caused by variations in reporting standards, safety procedures, or other elements affecting the frequency of accidents.

According to our model, the hazy peak in $\lambda_2$ points to a region that might need more research or data to fully grasp the subtleties affecting the event rates during this time. Furthermore, it implies that compared to other segments with more distinct peaks, predictions made for this one might not be as accurate.

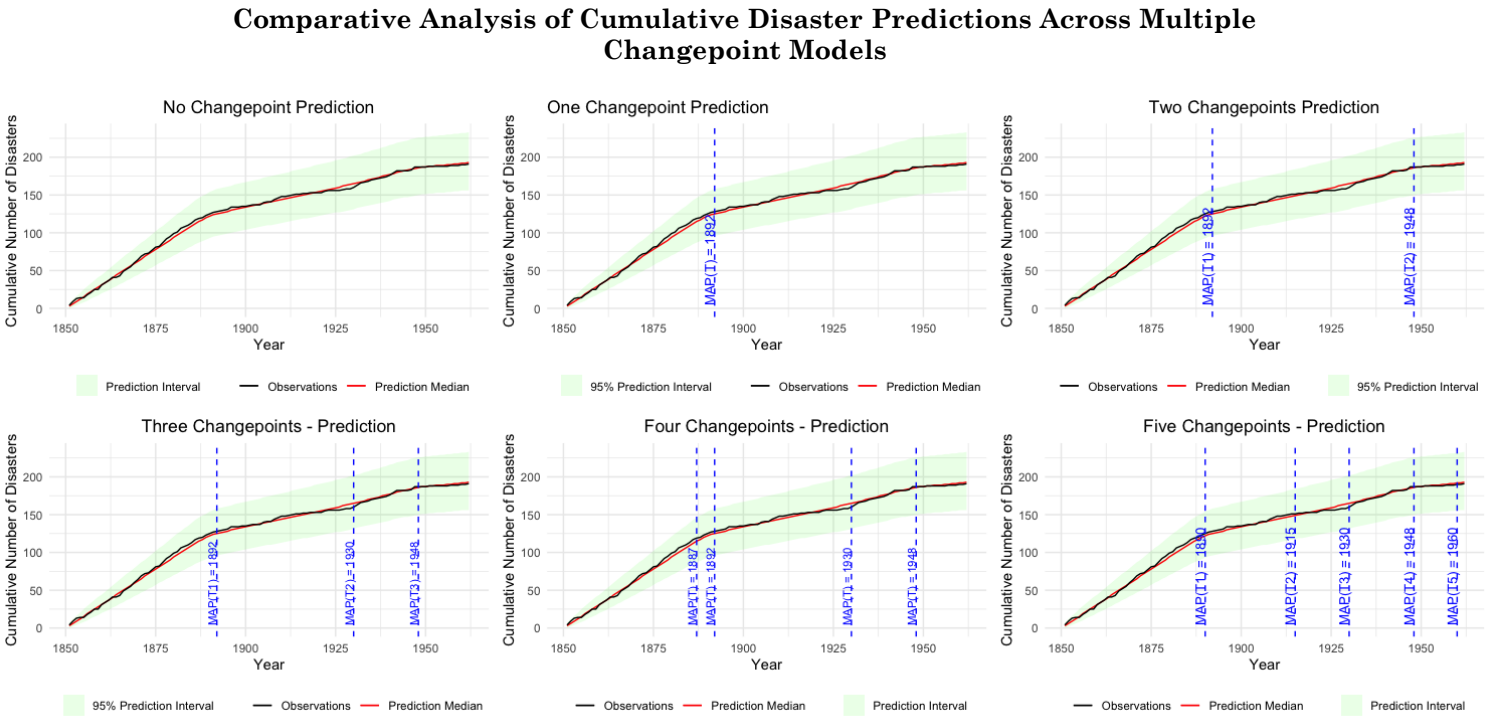


**Figure 12:** This series of plots illustrates the cumulative number of disasters predicted by Bayesian changepoint models with varying numbers of changepoints (from zero to five). Each subplot represents a different model, with the solid black line depicting the actual observed cumulative disasters, the red line indicating the median of the predicted values, and the green shaded area denoting the 95% prediction interval. Dashed blue lines mark the most probable changepoints (MAP estimates) identified within each model. These visualizations collectively demonstrate the increasing fit of the models to historical data as more changepoints are introduced, while also highlighting the potential for overfitting as the model complexity grows.

There is an obvious discrepancy between the observed data and the baseline model without changepoints, which assumes a constant rate of disasters over time. In particular, it overestimates later occurrences and underestimates early ones, indicating a large misfit. Consequently, this model's predictive power is limited because it is unable to account for variations in disaster rates that may arise from changing external factors.

A single changepoint greatly improves the model's fit to historical data, as shown by the median prediction line adhering more closely to trends that have been seen. According to the Maximum A Posteriori (MAP) estimation, the disaster rate underwent a significant change that the model could now account for around the year 1892. However, despite this improvement, the model's flexibility remains limited to the historical shift it has adapted to and may not capture future variations effectively.

A model with better historical fit is obtained by adding a second changepoint, suggesting that the disaster rate has experienced two significant regime changes. These are identified by MAP estimates for the years 1892 and 1948, which may be indicative of important historical occurrences or legislative modifications that affect the frequency of disasters. Although the two-changepoint model provides more dynamism, it can still miss some changes, especially if future shifts don't follow well-established historical patterns.

The model fits the data much more precisely with three changepoints, indicating that it can accommodate more variability in the disaster rate. A more detailed explanation of the changes in disaster occurrences can be found at the MAP changepoints in 1892, 1930, and 1948, which most

likely correspond to a variety of outside factors. This model is positioned to deliver more accurate short-term predictions due to its detailed historical representation.

A more precise fit can be seen in the four-changepoint model, which closely matches the total number of disasters that have been recorded. The model appears to capture a number of historical transitions, according to MAP changepoints at 1897, 1982, 1930, and 1948. This provides a thorough framework for comprehending the factors that influence disaster events. But this level of detail also increases the risk of overfitting, in which the model is too closely adjusted to past data, potentially undermining the accuracy of future projections.

The most intricate model, with five changepoints, shows the closest fit to the historical data, suggesting the most thorough identification of trends. There is a chance that the MAP changepoints at 1890, 1915, 1930, 1948, and 1960 will explain both large and small variations in the disaster rate. Yet, the heightened risk of overfitting with this model suggests that while it mirrors historical data effectively, its generalizability to new, unseen conditions is uncertain.

The comparison of changepoint models highlights the need to strike a careful balance between overfitting risk and model complexity. While models with a higher number of changepoints may provide a better historical fit, there is no guarantee that these models will accurately predict future events. In disaster management and risk assessment, choosing a model that fits historical data well and maintains predictive reliability for future conditions is essential for making informed decisions. In addition to encouraging more research, the search for the ideal model may lead to the addition of other predictive factors, which could improve the model's predictive range and fit.

## Model Comparison

**Table 10: Marginal Likelihood for Changepoint Models**

| Model | Marginal Likelihood |
|---|---|
| One Changepoint Model | 2.471304e-77 |
| Two Changepoint Model | 5.337402e-77 |
| Three Changepoint Model | 6.735974e-77 |
| Four Changepoint Model | 7.523818e-77 |
| Five Changepoint Model | 9.178468e-77 |

**Table 10**: The computed likelihoods for changepoint models ranging from one to five changepoints. These likelihoods quantify the probability of the data given each model, with higher values suggesting a better fit of the model to the observed data. The models are compared based on their ability to explain the variability in the data with different numbers of changepoints, which are locations in the data sequence where the statistical properties change.

The marginal likelihood is an important metric for model comparison in the Bayesian framework. It gives an indication of how well a model integrates over all possible parameter values to explain the observed data. An increasing trend in marginal likelihoods is revealed by our analysis across a variety of changepoint models as we add more changepoints, indicating a better fit to the historical data on coal mining disasters.

The plot (Figure 13) illustrates the marginal likelihoods for every model and highlights the law of diminishing returns, which is a crucial factor in model selection. The rate of increase in marginal likelihoods decreases as more changepoints are added, even though this indicates a better fit. The marginal likelihood may decrease with additional changepoint additions after it reaches a peak. Computational penalties resulting from model complexity are partially to blame for this phenomenon.

We find a plateau and then possible decrease in marginal likelihood beyond five changepoints, indicating that the model might be fitting noise in the data rather than meaningful patterns. Computational penalties are applicable in this situation. These penalties are part of the built-in Bayesian model comparison process, which favours simpler models unless they offer a significantly better fit. These penalties are not arbitrary.

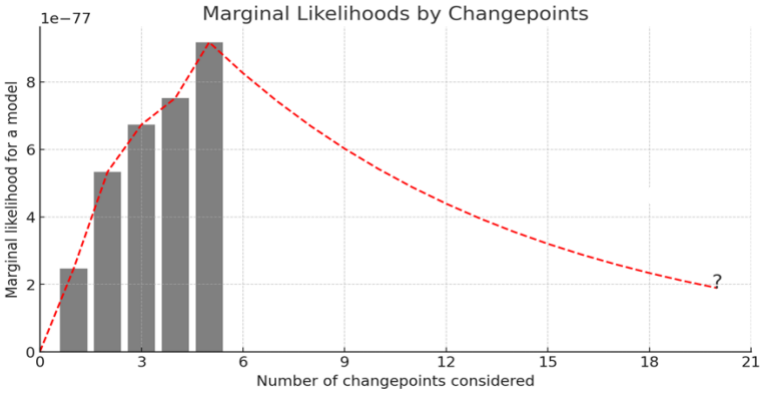


**Figure 13:** The marginal likelihoods for models with varying numbers of changepoints, aiding in the selection of the most suitable model to explain the historical coal mining disaster data.

When the model becomes too complex, it requires more computational resources to explore an increasingly vast parameter space. With too many changepoints, the model may begin to interpret random fluctuations as significant changes, leading to overfitting. The additional computational burden is not just a matter of longer run times but also reflects the increasing difficulty in achieving convergence and finding a global optimum in the parameter space.

Thus, while our analysis does indicate that models with four or five changepoints are most favored statistically, we must also consider computational efficiency and the risk of overfitting. The optimal model is one that achieves a balance, capturing the essential shifts in the data while remaining computationally manageable and interpretable. This optimal point is often found at the 'knee' of the curve, where adding further changepoints yields minimal improvements in fit relative to the increase in complexity.

Finally, the choice of final model should be based on a careful balance between simplicity and fit quality, making sure that the model is computationally feasible, robust, and comprehensible. Choosing a model that is both simple enough to prevent overfitting and retain practical usefulness and complex enough to capture the underlying structure of the data is the aim.

## Conclusion and Future Work

To identify the latent structure in the data, we used Bayesian changepoint models in our study of the temporal dynamics of coal mining disasters. The use of models with one to five changepoints revealed discrete time periods with varying rates of disasters. Our method's statistical soundness was supported by the computation of marginal likelihoods, which provided a numerical comparison of the relative strengths of each model.

We determined that adding multiple changepoints greatly improved our models' fit to the historical data through careful model comparison. More specifically, models with four or five changepoints were statistically better than simpler models, capturing subtle changes in the frequency of disasters that the simpler models missed. However, we must admit that there are diminishing returns to increasing complexity. The likelihood of fitting to random fluctuations rather than actual changes in the underlying process rises with the number of changepoints used. This could result in models that are more difficult to understand and less generalizable.

**Future Work: A Detailed Prospectus**

As we progress with our Bayesian changepoint analysis, our commitment to methodological improvement and computational efficiency remains steadfast. Here are our key strategies for future developments:

1. Algorithmic Enhancements:
    - **Transition to Gibbs Sampling:** Transitioning to Gibbs sampling promises improvements in sampling efficiency and parameter estimation accuracy, particularly advantageous for our complex, multi-changepoint models such as:
        - Improved Convergence: Gibbs sampling is renowned for faster convergence in complex models, such as ours with multiple changepoints, due to its strategy of updating parameters sequentially from their conditional distributions.

    - Ease of Implementation: Unlike MH, Gibbs sampling does not require the careful design of proposal distributions, which simplifies the algorithm's implementation and reduces the manual tuning process.
    - Sampling Effectiveness: by drawing samples directly from the conditional distributions, Gibbs sampling often achieves higher efficiency, which is beneficial in models with a large number of parameters.
- **Hamiltonian Monte Carlo Integration:** Incorporating Hamiltonian Monte Carlo methods to refine the exploration of the parameter space, reducing autocorrelation and improving convergence times. The adoption of Hamiltonian Monte Carlo (HMC) methods promises a leap forward in our sampling capabilities:
    - Enhanced Exploration: HMC utilizes derivatives of the posterior distribution to inform the sampling process, allowing for a more intelligent exploration of the parameter space.
    - Diminished Autocorrelation: By integrating over potential energy landscapes, HMC reduces random walk behaviour and autocorrelation, leading to more independent samples.
    - Accelerated Convergence: The method provides a mechanism for efficiently traversing high-dimensional spaces, potentially yielding faster convergence, and reducing the need for extensive iterations.
    - Adaptability and Scalability: HMC's adaptability to complex distributions makes it suitable for high-dimensional models, a characteristic beneficial for scaling to more intricate datasets.

2. **Model Complexity and Hierarchies:** Implementing hierarchical models to effectively capture the shared information and dependencies across different temporal or spatial segments of the data, mitigating overfitting concerns.

3. **Computational Advances:** Leveraging Approximate Bayesian Computation (ABC): Employing ABC to circumvent intractable likelihoods, facilitating a balance between computational feasibility and statistical detail.

4. **Exploiting Parallel Computing:** Utilizing parallel computing to accelerate the processing of complex models and large datasets.

5. **Model Robustness and Sensitivity Analysis:** Conducting thorough robustness checks to ensure our models' findings are reliable under various assumptions.

Our objective is to enhance the convergence behaviour of our models through the incorporation of Gibbs sampling into our set of techniques, thereby yielding a more sophisticated and perceptive data analysis. This refined process is a component of a broader endeavour to improve the accuracy and comprehensibility of our results, guaranteeing that our models maintain their statistical correctness and computational ease.

## References


Adams, R. P., & MacKay, D. J. C. (2007). Bayesian Online Changepoint Detection. arXiv:0710.3742.

Fearnhead, P. (2006). Exact and efficient Bayesian inference for multiple changepoint problems. Statistics and computing, 16(2), 203-213.

Taboga, M. (2021). "Metropolis-Hastings algorithm", Lectures on probability theory and mathematical statistics. Kindle Direct Publishing. Online appendix. https://www.statlect.com/fundamentals-of-statistics/Metropolis-Hastings-algorithm.

Toptal. (n.d.). Metropolis-Hastings and Bayesian Inference. Retrieved from https://www.toptal.com/algorithms/metropolis-hastings-bayesian-inference`` 【oaicite:0】``&#8203.

Unicsoft. (n.d.). Change Point Detection: Definition, Examples, and Types. Retrieved from https://unicsoft.com/blog/change-point-detection-definition-examples-and-types/.

Eckley, I.A., Fearnhead, P., & Killick, R. (2011). Analysis of changepoint models. In Bayesian Time Series Models (Cambridge Books Online). Cambridge University Press. Retrieved from https://www.cambridge.org/core/books/bayesian-time-series-models/analysis-of-changepoint-models/CBED86FD9B97C2EE8D33C156A0AB8B0E.

TensorFlow Probability Development Team. (n.d.). Multiple changepoint detection and Bayesian model selection. TensorFlow Probability. Retrieved from https://www.tensorflow.org/probability/examples/Multiple_changepoint_detection_and_Bayesian_model_selection.

Pamungkas, A. Bayesian Change Point Detection in Coal Mining Disaster Data. Medium. Retrieved from https://medium.com/@anggerpamungkas/change-point-detection-1b2e588fe6b0.