Universitat de Barcelona

Facultat de Matemàtiques i Informàtica

Fundamental Principles of Data Science

Bayesian Statistics Proyect

Bayesian Switchpoint analysis

Adriana Alvaro, Madison Chester,
Dafni Tziakouri, Arturo Fredes

2023-2024

Contents

1 Abstract

This project is conducted under the umbrella of the Bayesian Statistics Final Project for the Master's in Fundamental Principles of Data Science at the Universitat de Barcelona. The focus is on Bayesian Switchpoint Analysis. A historical and theoretical background of Bayesian Inference is provided, including definitions and an illustrative example for better understanding. This example is solved using both analytical and numerical methods. Probabilistic programming is employed for the experiments, which necessitates a theoretical explanation of this approach. Finally, experiments applying these theoretical tools are presented, specifically examining changes in the amplitude of a wave and the value of Bitcoin.

2 Introduction

In this work, Bayesian Switchpoint Analysis will be presented. This technique consists of Bayesian inference used to estimate changes that have occurred in the evolution of a stochastic process, thereby aiding in understanding its behavior. The method originated in 1950 with the aim of detecting failures in industrial processes. Today, it represents an active area of research with applications extending beyond quality control in industry, facilitated by advancements in computing capabilities.

In addition to introducing the concept and potential of this versatile technique, the demonstration will show how the Bayesian approach is more intuitive when modeling observed data. From a Bayesian statistical perspective, a time series can be understood as a realization of a stochastic process defined by certain parameters. This allows for the measurement of how good the hypothesis is, meaning how likely it is to obtain the observed data series, as opposed to a classical or frequentist approach where this probability might be assumed as the frequency of observed events over time.

In the technique to be discussed, specific interest lies in identifying the parameters that indicate changes in the data generation process. The belief about the values of these changes will be represented in the form of a probability distribution. This distribution, referred to as prior probability, encodes how the data was generated and marks a departure from the classical approach.

3 Bayesian Switchpoint Analysis:

3.1 Information Theory and Bayesian Inference

The scientific advancements during World War II led to significant progress, including the development of the Monte Carlo method for studying stochastic processes. In 1957, Jaynes proposed interpreting statistical physics through information theory, as established by Shannon, to infer a probability distribution consistent with data.

Bayesian inference, aligned with Jaynes's approach, updates this probability distribution with observational evidence. A time series of data, viewed as a realization of a stochastic process, forms the basis for modeling. The likelihood, representing the probability of obtaining the observed data, measures the model's accuracy.

While classical statistics view probability as the long-term frequency of events, Bayesian statistics interpret it as a belief, updated with new evidence. This prior probability, integrated into Bayes' Theorem, allows Bayesian methods to consider evidence, unlike classical methods which treat parameters as fixed but unknown.

Given a set of N observations $d = \{d_1, \ldots, d_N\}$, the goal is to determine the parameters Θ that best fit the data by maximizing the likelihood $P(d|\Theta)$. Bayes' Theorem,

$$P(\Theta|d) = \frac{P(d|\Theta)P(\Theta)}{P(d)}, \tag{1}$$

resolves the challenge of evaluating the probability of parameters given the data, incorporating the prior probability $P(\Theta)$. This theorem, formulated by Thomas Bayes and later refined by Laplace, underpins Bayesian inference, addressing the probability of parameter sets based on observed data.

3.1.1 Prior:

Prior probability, in Bayesian statistics, is the probability of an event before new data is collected. This is the best rational assessment of the probability of an outcome based on the current knowledge before an experiment is performed.

The prior probability of an event will be revised as new data or information becomes available, to produce a more accurate measure of a potential outcome. That revised probability becomes the posterior probability and is calculated using Bayes' theorem. Prior probability represents what is originally believed before new evidence is introduced.

3.1.2 Posterior:

A posterior probability, in Bayesian statistics, is the revised or updated probability of an event occurring after taking into consideration new information. The posterior probability is calculated by updating the prior probability using Bayes' theorem. In statistical terms, the posterior probability is the probability of event A occurring given that event B has occurred.

Posterior probability distributions should be a better reflection of the underlying truth of a data generating process than the prior probability since the posterior included more information. A posterior probability can subsequently become a prior for a new updated posterior probability as new information arises and is incorporated into the analysis.

3.1.3 Likelihood:

The likelihood function $L(\theta \mid x)$ is defined as a function of θ indexed by the realization x of a random variable with density $f(x \mid \theta)$:

$$L : \Theta \to \mathbb{R} \quad \theta \mapsto f(x \mid \theta) \tag{2}$$

3.2 Application Example

The Switchpoint Analysis technique uses the method of maximizing the likelihood. It is assumed that a stochastic process changes at some point, and the goal is to determine this instant, denoted by a probability distribution τ. Bayesian inference focuses on the parameter τ from the set Θ, which indicates the change in data behavior. The objective is to obtain the posterior distribution of τ: $P(\tau \mid d)$.

Consider a simple example where data is generated from a Gaussian distribution with a mean μ and variance σ^2. At a certain time τ, the mean μ changes. Let μ_1 be the mean before the change and μ_2 be the mean after the change. Each data point d_i in the observed set d is generated as:

$$d_i = \begin{cases} N(\mu_1, \sigma) & \text{if } t < \tau \\ N(\mu_2, \sigma) & \text{if } t \geq \tau \end{cases} \tag{3}$$

Thus, the parameters are $\Theta = \{\mu_1, \mu_2, \sigma, \tau\}$ and the observed data d_i. The objective is to infer the parameters Θ from the observations d.

3.2.1 Analytics solution

In this section, an analytical solution of the algorithm in question will be obtained, taking Bayes' Theorem (1) as the starting point and based on Chapter 5 of the book *Numerical Bayesian Methods Applied to Signal Processing* by Joseph O. Ruanaidh and William Fitzgerald.

If we apply Bayes' Theorem (1) to this model (3):

$$P(\mu_1, \mu_2, \sigma, \tau \mid d) = \frac{P(d \mid \mu_1, \mu_2, \sigma, \tau) P(\mu_1, \mu_2, \sigma, \tau)}{P(d)} \tag{4}$$

On one hand, $P(\mu_1, \mu_2, \sigma, \tau \mid d)$ is the posterior distribution, meaning the updated probability after considering the observed data and evidence. Secondly, $P(d \mid \mu_1, \mu_2, \sigma, \tau)$ is the likelihood, indicating how probable it is to observe the data we have for a certain set of model parameters. Therefore, it is the distribution we aim to maximize. The prior probability $P(\mu_1, \mu_2, \sigma, \tau)$ encodes the distribution of each model parameter before observing the data. Finally, $P(d)$ is the evidence, a measure of how good our model is. For the purpose of this section, we will consider this term simply as a multiplicative factor.

Given the lack of specific prior information on how the data was generated, a common initial choice for the prior distributions are the non-informative ones. These distributions, also called diffuse priors, are flat compared to the likelihood and reflect significant uncertainty about the parameter information. One of the most popular methods for obtaining such distributions is Jeffreys' Rule [2]. These types of distributions can result in improper priors, which do not integrate to 1. Assuming that μ_1, μ_2, σ, and τ are independent parameters:

$$\begin{aligned}P(\mu_1) &= k_1,\\ P(\mu_2) &= k_2,\\ P(\log\sigma) &= k_3 \implies P(\sigma) = \frac{1}{\sigma},\\ P(\tau) &= k_4,\end{aligned}$$

where k_1, k_2, k_3, and k_4 are unspecified constants.

Our objective in this analysis is to obtain the posterior distribution $P(\tau \mid d)$. To do this, we will marginalize the posterior distribution $P(\mu_1, \mu_2, \sigma, \tau \mid d)$ with respect to the parameters that are not of interest (all except τ). Marginalizing simply involves integrating over the parameters we are not interested in, leaving us with a single dimension of the distribution.

If we consider:

$$P(\mu_1, \mu_2, \sigma, \tau) = P(\mu_1)P(\mu_2)P(\sigma)P(\tau) = k_1 k_2 k_4 \frac{1}{\sigma}$$

and introduce it into (3), we get:

$$P(\tau \mid d) \propto \int_0^\infty d\sigma \int_{-\infty}^\infty d\mu_1 \int_{-\infty}^\infty d\mu_2 \frac{P(d \mid \mu_1, \mu_2, \sigma, \tau)}{\sigma}. \tag{5}$$

To obtain the distribution of τ, we need the likelihood function $P(d \mid \mu_1, \mu_2, \sigma, \tau)$. According to the defined model (2), we assume that from 1 to τ the data is modeled by a Gaussian distribution with mean μ_1 and variance σ. On the other hand, from τ onwards, they are modeled by a Gaussian distribution with mean μ_2 and the same variance. Thus:

$$\begin{aligned}P(d \mid \mu_1, \mu_2, \sigma, \tau) &= \prod_{i=1}^{\tau} P(d_i \mid \mu_1, \sigma) \prod_{j=\tau+1}^{N} P(d_j \mid \mu_2, \sigma)\\ &\propto (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{\tau}(d_i^2 + \mu_1^2 - 2\mu_1 d_i) + \sum_{j=\tau+1}^{N}(d_j^2 + \mu_2^2 - 2\mu_2 d_j)\right]\right).\end{aligned} \tag{6}$$

Using the following identity,

$$\int_{-\infty}^{\infty} \exp\Big(-ax^2 - bx - c\Big)\, dx = \sqrt{\frac{\pi}{a}} \exp\left(\frac{b^2}{4a} - c\right), \tag{7}$$

we will integrate with respect to the variable μ_1:

$$\begin{aligned}P(\mu_2, \sigma, \tau \mid d) &= \int_{-\infty}^{\infty} P(\mu_1, \mu_2, \sigma, \tau \mid d)\, d\mu_1\\ &\propto \left(\frac{\tau}{2\pi}\right)^{1/2} (2\pi\sigma^2)^{-N/2} \exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{\tau} d_i^2 - \frac{\left(\sum_{i=1}^{\tau} d_i\right)^2}{\tau} - \sum_{j=\tau+1}^{N}(d_j - \mu_2)^2\right]\right).\end{aligned} \tag{8}$$

And with respect to μ_2:

$$
\begin{aligned}
P(\sigma, \tau \mid d) &= \int_{-\infty}^{\infty} P(\mu_2, \sigma, \tau \mid d)\, d\mu_2 \\
&\propto (2\pi\sigma^2)^{-(N-1)/2} \left(\tau(N-\tau)\right)^{1/2} \exp\left(-\frac{1}{2\sigma^2} \left[\sum_{i=1}^{\tau} d_i^2 - \frac{\left(\sum_{i=1}^{\tau} d_i\right)^2}{\tau} - \frac{\left(\sum_{j=\tau+1}^{N} d_j\right)^2}{N-\tau} \right] \right).
\end{aligned}
\tag{9}
$$

Finally, using the identity

$$
\int_0^{\infty} x^{\alpha-1} \exp(-Qx)\, dx = \frac{\Gamma(\alpha)}{Q^{\alpha}}, \tag{10}
$$

where $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} \exp(-x)\, dx$, is the Gamma function, we will integrate with respect to σ. We finally obtain the desired distribution for the switchpoint τ:

$$
P(\tau \mid d) = \int_0^{\infty} P(\sigma, \tau \mid d)\, d\sigma \propto \frac{1}{\tau(N-\tau)} \left[\sum_{k=1}^{N} d_k^2 - \frac{\left(\sum_{i=1}^{\tau} d_i\right)^2}{\tau} - \frac{\left(\sum_{j=\tau+1}^{N} d_j\right)^2}{N-\tau} \right]^{-\frac{N-2}{2}}. \tag{11}
$$

We have obtained an expression for the posterior distribution of the parameter τ ($P(\tau \mid d)$), which does not depend on the other parameters μ_1, μ_2, and σ, as we have integrated with respect to them. Therefore, $P(\tau \mid d)$ depends exclusively on the observables d.

3.2.2 Numerical Solution of Example

The numerical solution is obtained using the Python package PyMC3. Prior to development, key aspects of the method are highlighted. Many problems use methods to approximate the likelihood distribution, often employing Markov Chain Monte Carlo (MCMC) methods to generate random samples from this distribution. These chains converge to a stationary solution corresponding to the posterior distribution. Advancements in computing have facilitated the implementation of this technique, which was previously computationally expensive.

Correct choice of prior probabilities accelerates convergence, but describing an efficient algorithm without manual data inspection remains challenging. This is akin to supervised learning, requiring adjustment of initial parameters for a general solution.

First, an example compares equation (11), obtained analytically, with the one reconstructed by PyMC3 for a single switchpoint τ. Random data generation illustrates the inference of the target distribution. Subsequently, a dataset with two switchpoints, τ_1 and τ_2, is used.

A model is generated where data originates from a normal distribution with a variance of $\sigma = 50$ and a mean $\mu_1 = 1000$, changing at $\tau = 3000$ to $\mu_2 = 1040$:

$$d_i = \begin{cases} N(1000, 50) & \text{for } t < 3000 \\ N(1040, 50) & \text{for } t \geq 3000 \end{cases} \tag{12}$$

Based on this model, we will generate a set of random data, which is represented in Figure 1.

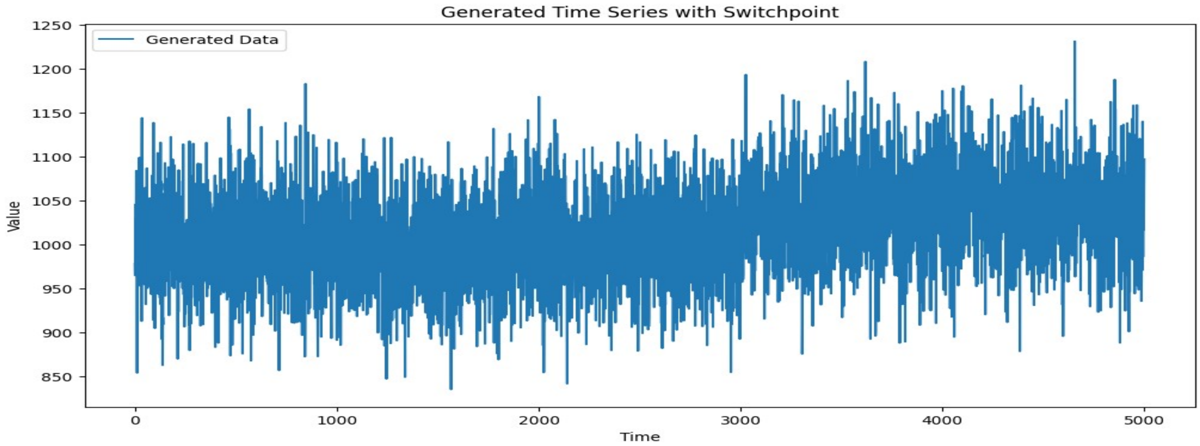


Figure 1: *Time series generated from the model.*

At first glance, it may not be apparent, but as described in the model, there is a switchpoint at $\tau = 3000$. Implementing the model with the aforementioned characteristics, but without pre-set values for μ_1, μ_2, σ, and τ, the distributions to which the generated samples have converged are obtained, as shown in Figure 2. These figures exclude the initial simulated values (burn-in iterations) as they are discarded for not being in the stationary state.

The results of the Monte Carlo iterations are shown in the right column. The sample values show little variation, indicating algorithm convergence to a stationary solution. The histogram in the left column shows this stationary solution as the posterior probability $P(\mu_1, \mu_2, \sigma, \tau \mid d)$. Each row corresponds to the marginal distributions $P(\tau)$, $P(\mu_1)$, $P(\mu_2)$, and $P(\sigma)$, respectively.

To ensure convergence, the analysis was repeated with different initial conditions. Figure 2 (left) shows the distribution obtained for each independent Monte Carlo chain. When these chains converge to a stationary solution, they are very close, indicating good convergence. The spectrum of the samples generated for each parameter is shown on the right, indicating the algorithm has explored the entire possible range.

In Figure 3, the observed data is depicted with black points, highlighting the switchpoint at $\tau = 3000$. The figure also presents the posterior median as a blue line, accompanied by the 95% quantile intervals.

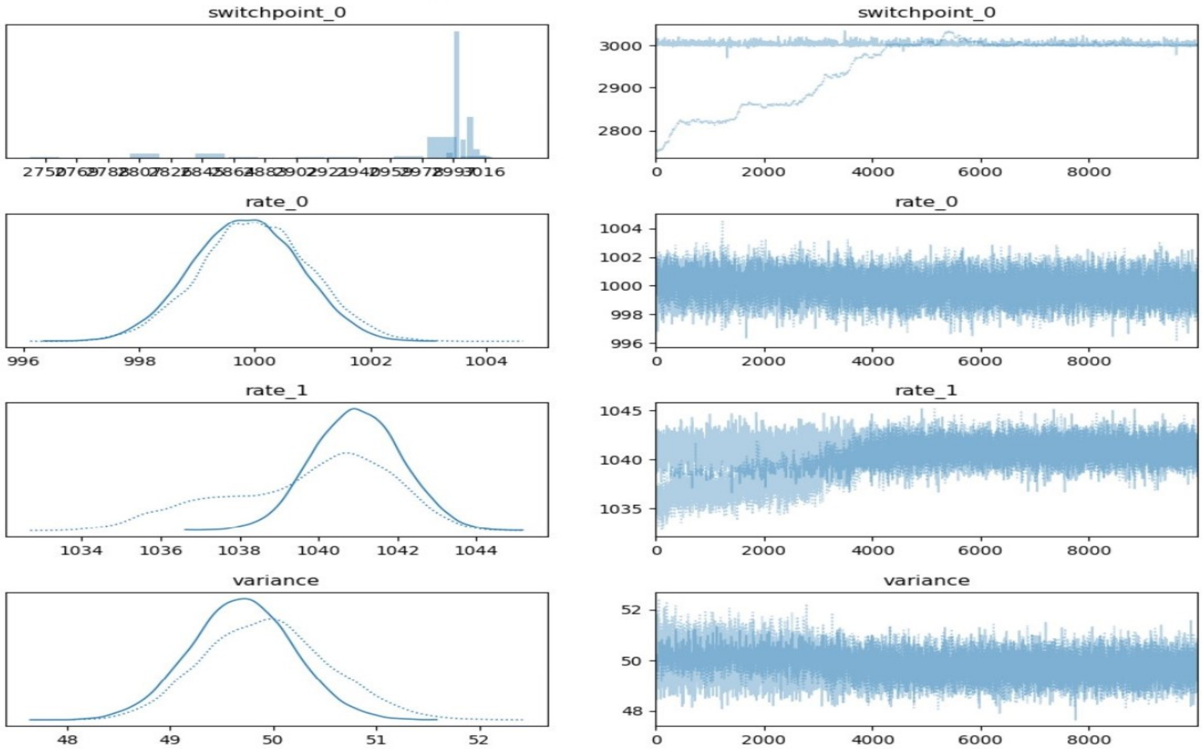


Figure 2: *Samples generated with different Monte Carlo chains. On the right, their values for each iteration. On the left, the corresponding histograms for each chain. In all parameters, these histograms are very close, indicating that the algorithm has correctly converged to a stationary solution, the posterior distribution.*

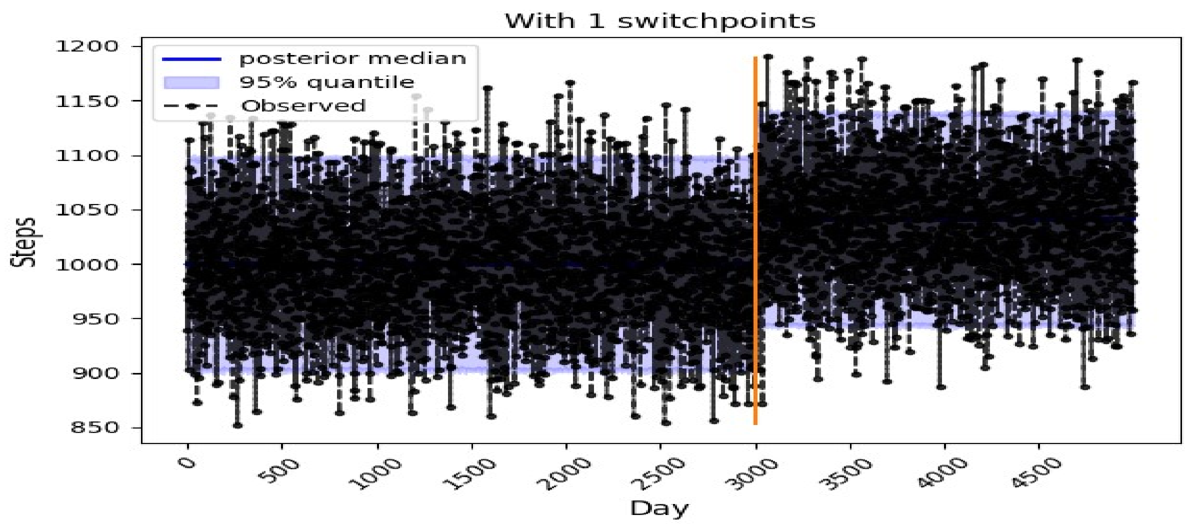


Figure 3: *Time series generated from the model with the switchpoint at* $\tau = 3000$.

4 Probabilistic Programming

Bayesian statistical methods have a drawback that has limited their use within the scientific community. Evaluating Bayesian evidence and, in general, the likelihood function can be very computationally expensive. The likelihood function is generally non-analytic, and its values can be studied using Monte Carlo methods. Currently, there are developed Bayesian analysis libraries, such as PyMC3, which its developer, Thomas Wiecki, calls the inference button. "These libraries are making these methods increasingly easier to implement and should be used more frequently." This is a message that the author of a 2020 reference paper [6] conveyed, particularly to the atomic physics community, where Bayesian methods had not been fully adopted.

A typical classical programming scheme would be to write a program, specify the values of its arguments, and evaluate the program to produce an output. On the other hand, modeling in statistics would follow an approach like the one illustrated in Figure 4. It would start with the output, which would be the observations or data (d), and a generative model of these data $P(\Theta, d)$ would be defined. Finally, inference techniques would be used to characterize the posterior distribution $P(\Theta|d)$ of the unknown parameters, given the observables.

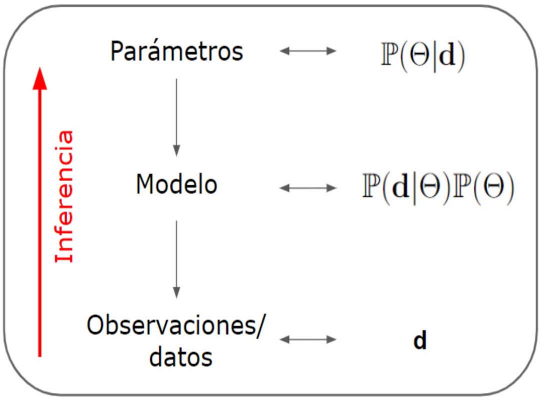


Figure 4: *Typical Statistical Inference Approach Scheme.*

On the other hand, probabilistic programming treats variables as probability distributions, unlike classical programming, which treats them as numbers. This adds some complexity because to display the value of a variable, one would need to draw random numbers from the distribution.

Bayesian statistics is conceptually simple; we have knowns and unknowns. With Bayes' Theorem, we condition one with the other, reducing the uncertainty of the unknowns. On the one hand, the knowns are the observables or data, treated as constants. On the other hand, the unknowns are the parameters, which, unlike in classical programming, are treated as distributions. That is, with Bayes' Theorem, we transform the prior probability $P(\Theta)$ to the posterior probability $P(\Theta|d)$.

This programming concept is called Probabilistic Programming (PP). It is probabilistic in the sense that we create probability models using distributions as model parameters. B. Cronin defines PP as:

"Unlike a traditional program, which only compiles in the forward direction, a probabilistic program compiles in both directions, forward and backward. Forward, to calculate the consequences of the assumptions, but also backward (from the data), to rule out impossible explanations. In practice, many probabilistic programming systems perform operations in both directions to effectively conclude the best explanations or models."

Probabilistic Programming offers an effective way to build and solve complex models, allowing us to focus more on model design, evaluation, and interpretation, and less on mathematical or computational details. It has enabled us to automate the inference process and make a clear separation between model creation and inference.

In this work, the models have been implemented using PyMC3, the Python library renowned for Bayesian analysis. The core of PyMC3 is written in Python, while the components with higher computational demands are written in NumPy and Theano.

Theano [3] is a library that enables the definition of expressions using generalized data vector structures, called tensors. Its notation is very similar to that of NumPy. However, whereas NumPy directly executes the calculations, Theano constructs a "computation graph," which maps out the operations to be performed. The operation itself is carried out when a Theano expression is evaluated.

Once this computation graph is constructed, mathematical optimizations can be performed, such as compiling it in C to enhance speed. PyMC3 consists of a collection of symbolic Theano expressions for various probability distributions that combine into a computation graph, representing the probabilities of the entire model. For the Switchpoint Analysis technique, the change in parameters in the distribution is marked by a Theano function with the appropriate condition: switch().

4.1 Numerical Implementation

In this section, the method used to implement the numerical solution in the applications developed in section 5 is described.

A general common scheme was followed in all cases. First, the data was managed and organized using the Pandas library in Python. Next, the model was defined with PyMC3, initializing the distributions of each involved parameter and the likelihood distribution. Finally, samples were generated using Markov Chain Monte Carlo (MCMC). The correct convergence of these samples to a stationary solution, representing the sought posterior distribution, was verified. This process concluded with plausible explanations that shape the obtained model.

5 Experiments

5.1 Changes in the amplitude of a wave

In this experiment the aim is to find the switchpoint where a wave changes its amplitude. The data is generated with some noise and a certain wave amplitude specified and then the parameters are changed to generate a wave with different amplitude. In fig. 5.

The priors chosen to create the model are:

- $\tau_1 \sim UniformDiscrete(0, 3000)$, $\tau_2 \sim UniformDiscrete(3000, 6000)$ for the switchpoints since there are 6000 steps and they could potentially be any of those. To ensure better convergence we introduced the prior knowledge of the approximate position of the switchpoint (first or second half of the data)

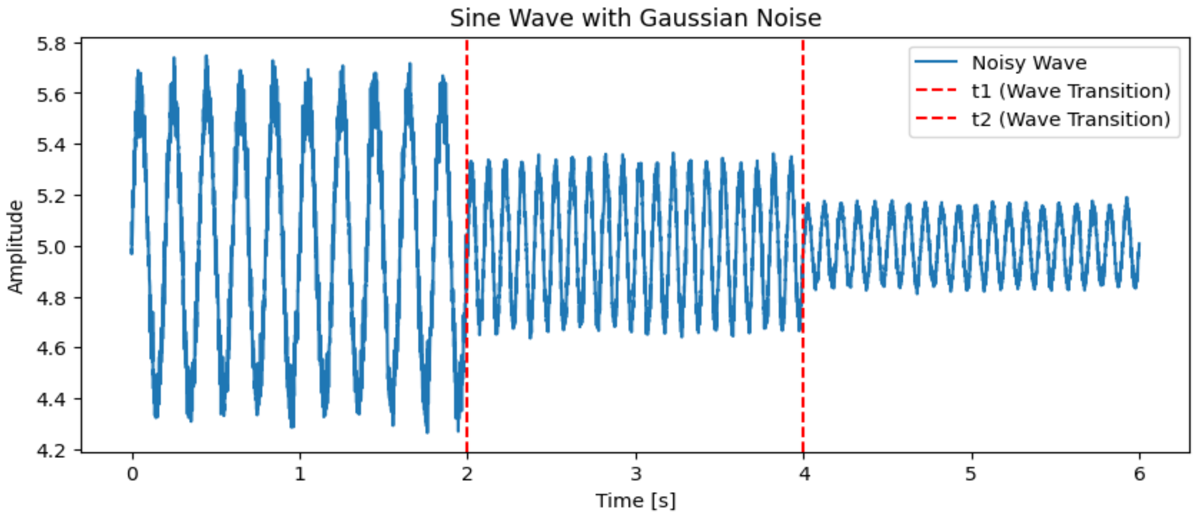


Figure 5: *Generated waves with different amplitudes. Source: own source.*

- The prior used for the data is a Normal distribution, and we will have two different distributions for σ since this parameter will be indicative of the distribution shift. For parameters, we also used a uniform distributions as non-informative priors.
- $\sigma_i \sim Uniform(0, 10)$
- $\mu \sim Uniform(1, 10)$

For each switchpoint a distribution is created, and for each interval a different distribution for σ. The MCMC algorithm is then used to sample and update the parameters, with the package PyMC3 in python. In figures 6, 7, 8, 9 and 10 can be seen the results of the MCMC. Plots on the left column show the posterior distributions estimated and plots on the right the results of the iterations. We observe that the switchpoint distributions are centered near 2000 and 4000 respectively, that the variance of the first wave (sigma 0) is bigger than sigma 1, and that the mean remains constant. In figure 11 the generated waves with the estimated parameters for the variances and the switchpoints.

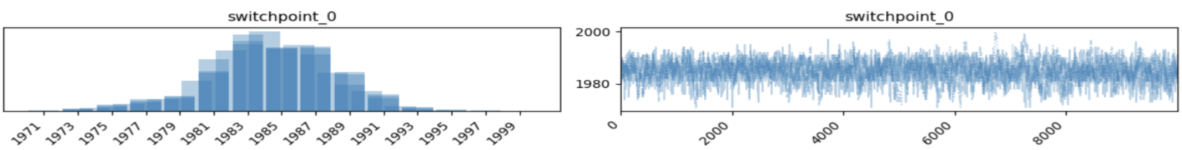


Figure 6: *Switchpoint 0 posterior distribution*

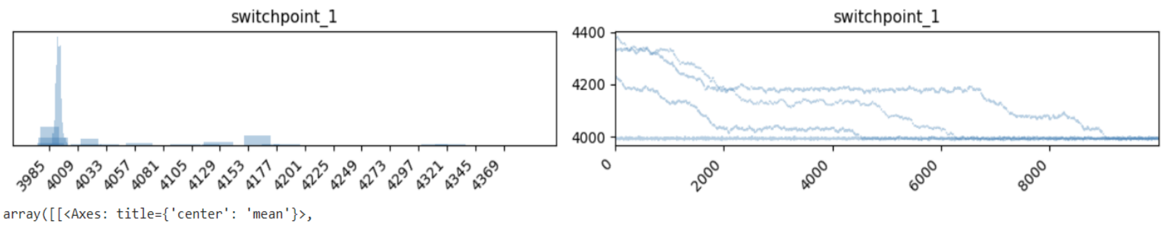


Figure 7: *Switchpoint 1 posterior distribution*

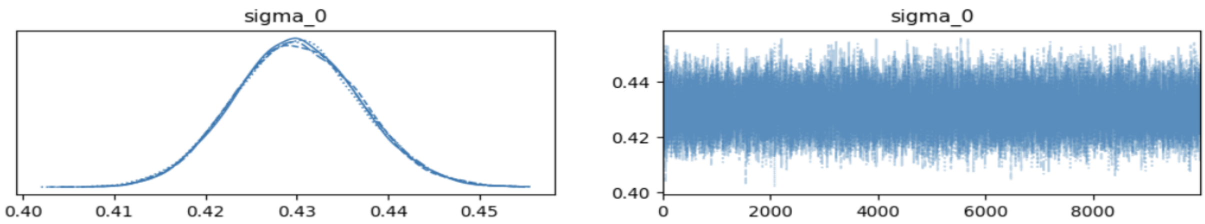


Figure 8: *Sigma 0 distribution*

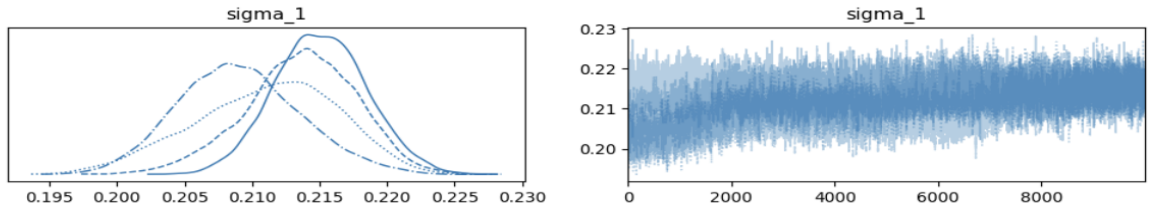


Figure 9: *Sigma 1 distribution*

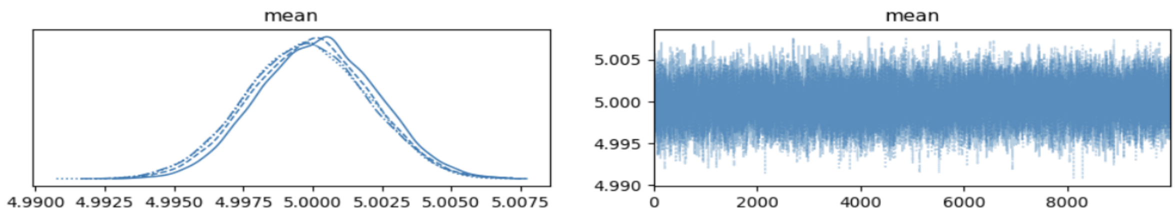


Figure 10: *Mean distribution*

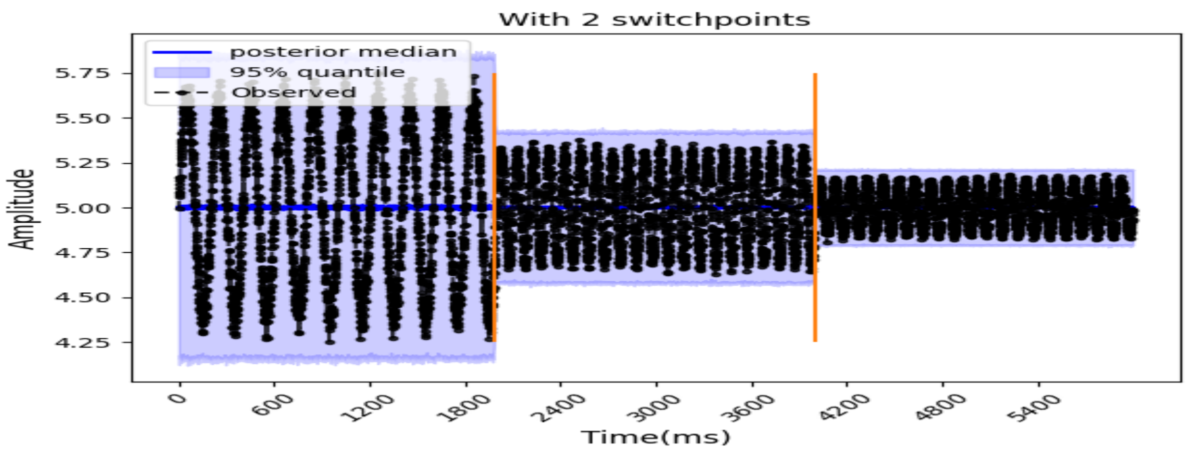


Figure 11: *Generated wave with estimated variances and switchpoints*

5.2 Bitcoin value

For the second experiment data from Bitcoin value along the first semester of 2021 was used. The idea was to try to find the switchpoints that corresponded to certain announcements concerning bitcoins done by the owner of Tesla Elon Musk. In figure 12 can be seen the evolution of the bitcoin value in the first semester of 2021 together with the timeline of the different announcements.

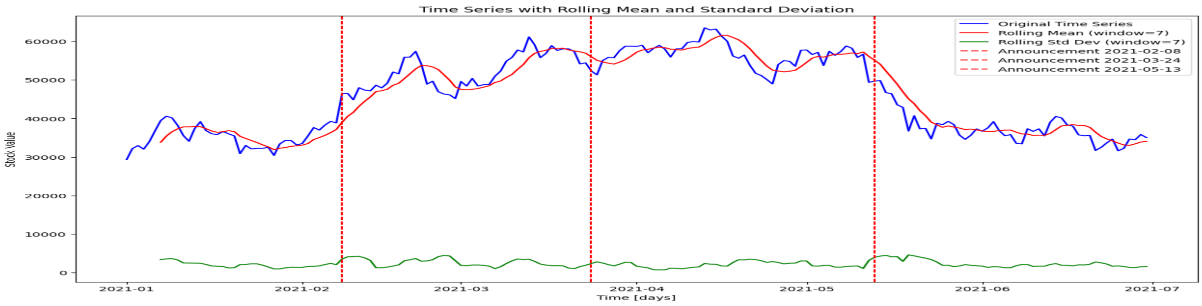
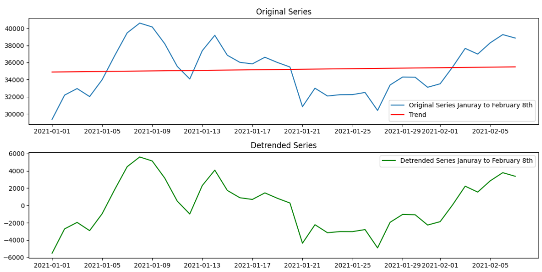


Figure 12: *Bitcoin value evolution first semester 2021*

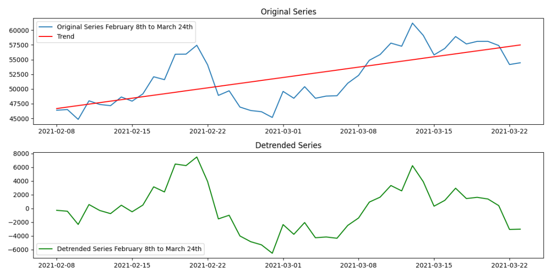
The first piece of news was in February 8th, in which Elon Musk reported a high value purchase of bitcoins [4]. The value of the coin raised considerably after it.

In March 24th Musk announced that purchasing Tesla cars with bitcoins was already possible [1]. But in May 13th he announced that bitcoins were not accepted anymore to purchase Tesla cars [5], and the value of the coin decreased.

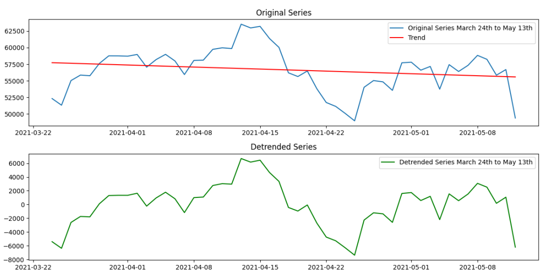
In figures 13a, 13b, 13c and 13d can be observed the 4 original time series (the ones we get when dividing by each timeline), a linear regression fitted with each time series (to see if there is any trend) and the detrended time series (the original series differencing the value of the trend and each point). It is observed that there seems to be an upward trend in 13b and a downward trend in 13d



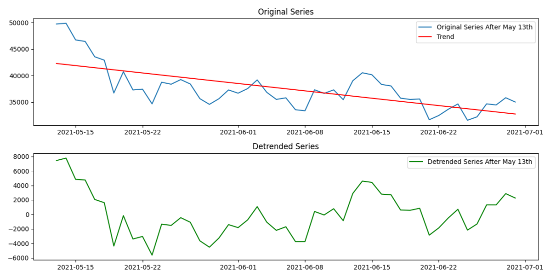
(a) *January to February 8th*



(b) *February 8th to March 23th*



(c) *March 23th to May 13th*



(d) *May 13th to July*

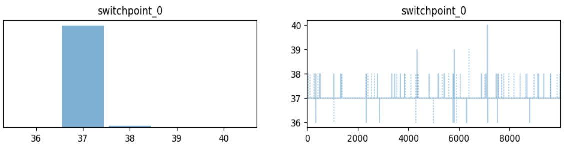
Figure 13: *Original and detrended time series from first semester 2021. Source: own source*

The switchpoint analysis has been done considering first just two possible switchpoints and then considering 3. The parameters to be estimated and their priors are:

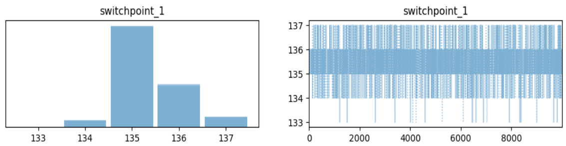
- $\tau_i \sim UniformDiscrete(0, ndays)$ for the switchpoints since any of the days can be a switchpoint. This time we did not introduce prior knowledge of the approximate position of the switchpoint.

- The prior used for the data was a negative binomial distribution, and again, we defined its parameters as uniform distributions.
- Rates: the mean of the data per partition (as many rates as switchpoints+1 considered). Uniform distributions.
- Variance: Gamma distribution shape parameter. As a uniform distribution and just one parameter for the whole data.

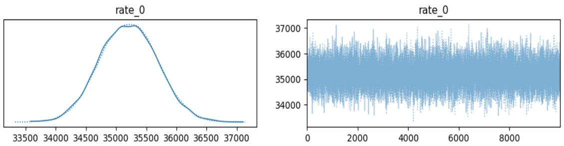
The analysis with two switchpoints seems to converge ton converge to the values that matches with the announcements of February 8th and May 13th (see figure 15).



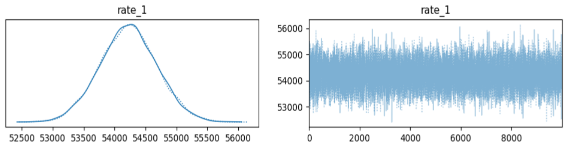
(a) *Switchpoint 0*



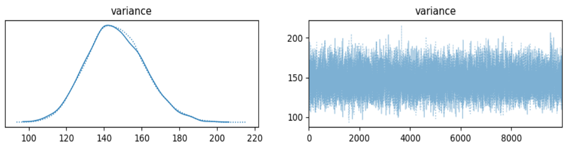
(b) *Switchpoint 1*



(c) *Rate 0*



(d) *Rate 1*



(e) *Variance*

Figure 14: *2 Switchpoint Analysis results*

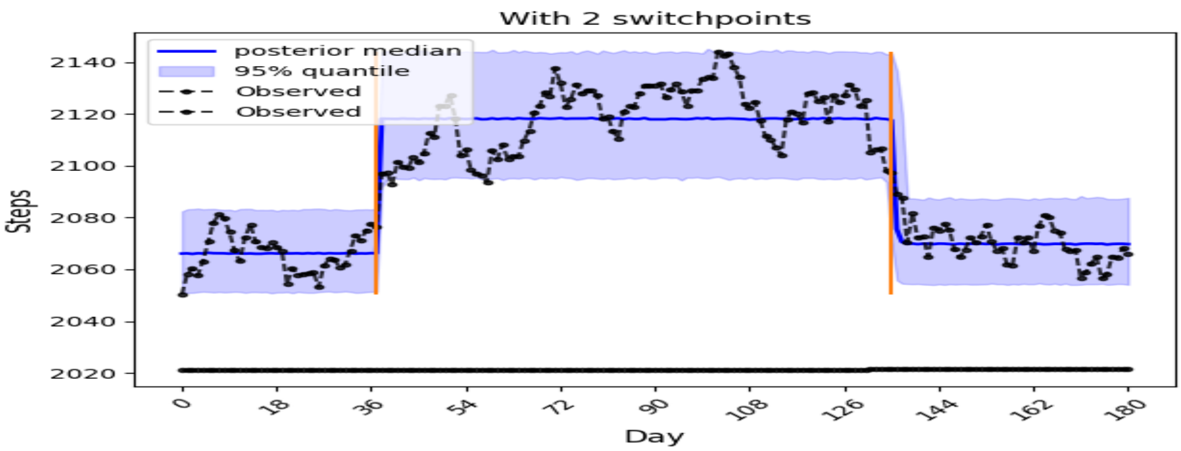
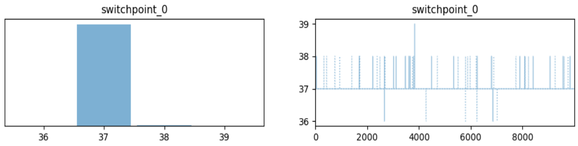
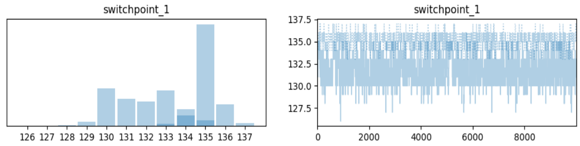


Figure 15: *Observed data and the posterior of the rate parameters.*

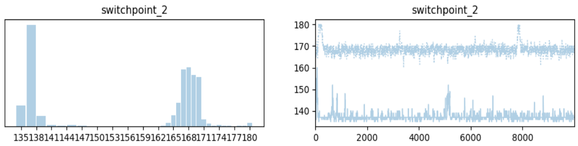
On the other hand, the analysis for 3 switchpoints, switchpoint 0 and 1 seem to converge to the same dates (February 8th and May 13th), but the third one does not match any previously known timeline. Furthermore, the one from March 24th does not seem to be found by the algorithm.



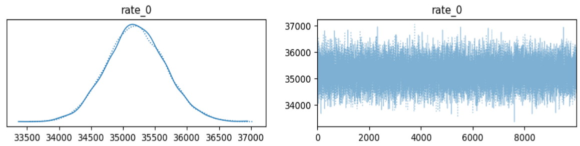
(a) *Switchpoint 0*



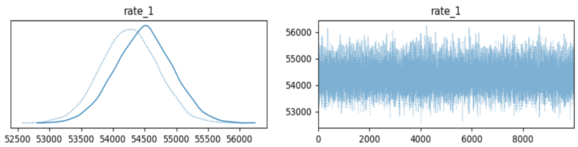
(b) *Switchpoint 1*



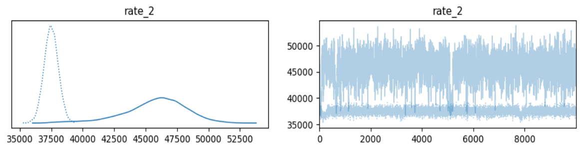
(c) *Switchpoint 2*



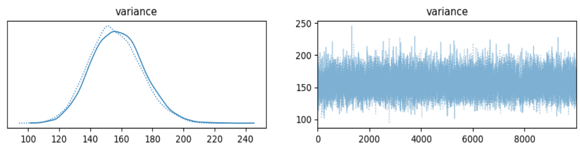
(d) *Rate 1*



(e) *Rate 1*



(f) *Rate 2*



(g) *Variance*

Figure 16: *3 Switchpoint Analysis results*

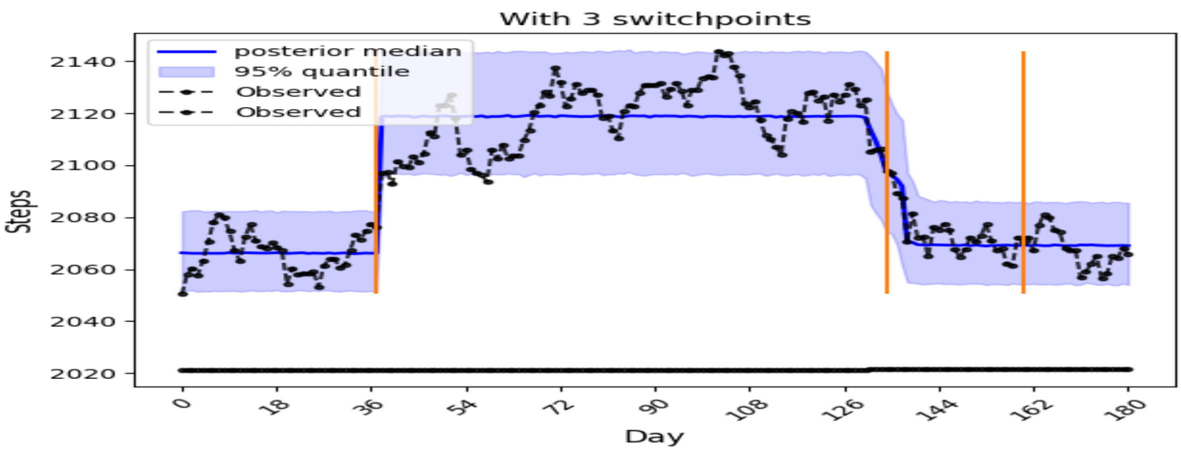


Figure 17: *Observed data and the posterior of the rate parameters for 3 switchpoint.*

In both analysis the variance distribution remains very similar. Regarding the rates, the bell of their distributions are centered at different values for the different partitions of the data, therefore indicating the change in the parameters of the data before and after the switchpoints.

6 Conclusions

In this work, we have introduced the fundamentals of the Bayesian Switchpoint Analysis technique and its wide-ranging applications for the analysis of time series. The Bayesian approach to the problem has allowed us to infer the posterior distributions of the model parameters that generate the data. We have previously defined this model in such a way that it fits the type of data, based on the information we had available as observers. We have focused on the posterior distribution of the parameter τ, which corresponds to the moment when a change occurred in the way the data are generated.

Bayesian Switchpoint Analysis shows to be a very good tool in handling uncertainties and incorporating prior knowledge for detecting changes in complex time series across various fields. Its adaptability and rigorous statistical framework enable accurate analysis and informed decision-making, highlighting its importance as a crucial analytical method.

References

[1] Computer Hoy. *Elon Musk anuncia que ya se puede comprar un Tesla con Bitcoin.* 2021. URL: https://computerhoy.com/noticias/motor/elon-musk-anuncia-ya-puede-comprar-tesla-bitcoin-835377.

[2] Harold Jeffreys. "An invariant form for the prior probability in estimation problems". In: (1946). URL: https://royalsocietypublishing.org/doi/10.1098/rspa.1946.0056.

[3] Christopher Fonnesbeck John Salvatier Thomas V. Wiecki. "Probabilistic programming in Python using PyMC3". In: (2016). URL: https://peerj.com/articles/cs-55/.

[4] BBC Mundo. *Bitcoin: por qué Tesla invirtió US*1.500*millonesenlacriptomoneda*. 2021. URL: https://www.bbc.com/mundo/noticias-55988372.

[5] BBC News Mundo. *Bitcoin y Elon Musk: el CEO de Tesla anuncia que ya no aceptarán la criptomoneda y esta sufre una fuerte caída.* 2021. URL: https://www.bbc.com/mundo/noticias-57096818.

[6] M. Trassinelli. "An introduction to Bayesian statistics for atomic physicists". In: (2020). URL: https://hal.science/hal-02885446/document.