**UNIVERSIDAD COMPLUTENSE DE MADRID**

**FACULTY OF PHYSICS**

ASTROPHYSICS DEPARTMENT


# BACHELOR FINAL THESIS

Code: FTA16

Analysis of astrophysical data using Bayesian statistics

*Análisis de datos astrofísicos usando Estadística Bayesiana*

Supervisor: Francisco Javier Gorgas García

**Ricardo Buitrago Ruiz**

Bachelor in Physics

June of 2021

**Determining the mass of a black hole from an upper bound observation, using a bayesian joint analysis of its host galaxy's variables.**

*Determinación de la masa de un agujero negro a partir de la medida de una cota superior, usando un análisis bayesiano conjunto de variables de su galaxia anfitriona.*


## Abstract:

In Astrophysics, sometimes it is not possible to calculate the mass of a black hole, it is only possible to provide an upper bound for it. In this thesis, we will find a probability distribution for the mass of such black holes. To do so, we will analyse seven variables of its host galaxy: the FUV-IR color, the NUV-IR color, the logarithm of the velocity dispersion, the logarithm of the effective radius, the logarithm of the K-band luminosity, the $3.6\mu m$-band absolute magnitude, and the logarithm of the mass in the K band. Instead of analysing them separately, we will introduce a novel probability distribution based on the product of the bayesian posteriors of each variable and inspired in the Bhattacharyya distance, that takes into account all the data to perform a joint analysis. This innovative joint distribution produces robust predictions on the black hole masses, that are not affected by the biases of individual linear regression predictions. Parallelly, we will develop a statistical model to incorporate the upper bound observations into our analysis, which we will incorporate into this joint fit. The calculations will be implemented in Stan.

## Resumen:

En Astrofísica, a veces no es posible calcular la masa de un agujero negro, sólo es posible calcular un límite superior para la misma. En este TFG, encontraremos una distribución de probabilidad para las masas de dichos agujeros negros. Para conseguirlo, analizaremos siete variables de sus galaxias anfitrionas: el color FUV-IR, el color NUV-IR, el logaritmo del radio efectivo, el logaritmo de la dispersión de velocidades, el logaritmo de la luminosidad en banda K, la magnitud absoluta en la bada $3.6\mu m$ y el logaritmo de la masa en la banda K. En vez de analizar estas variables por separado, introduciremos una nueva distribución de probabilidad basada en el producto de los posteriores bayesianos para cada variable e inspirada en la distancia de Bhattacharrya, que tiene en cuenta toda la información para realizar un análisis conjunto. Esta novedosa distribución conjunta produce predicciones robustas de la masa de los agujeros negros, que no están afectadas por los sesgos de las regresiones lineales individuales. Paralelamente, desarrollaremos un modelo estadístico para incorporar los límites superiores a nuestro anásis, que utilizaremos en el ajuste conjunto. Los cálculos se implementarán en Stan.

# Contents

# 1 Introduction

## 1.1 The role of Bayesian Statistics in Astrophysics.

Astrophysics tries to understand the universe beyond our planet. To do so, we must measure physical variables of far-away objects, yet every measurement that we make is not exact, it comes with an observation error. Furthermore, sometimes we can not measure the value of the physical variable itself, rather, we can only find some restriction to it (like an upper bound). That is to say, in nature there are *real* quantities that we want to measure, but we only have *observations* of such quantities, which are never exact. To deal with these limitations, we must work in a flexible framework that allow us to treat the *real* quantities like statistical parameters, and use our *observations* to estimate the probability distributions of such parameters. The discipline that allow us to do that is Bayesian Statistics.

In this thesis, we will denote our observations as $Z$ and the real quantities as $\Theta$. Both $Z$ and $\Theta$ are vectors of random variables and we must use our physical understanding of the problem to determine the joint, marginal and conditional distributions of $(Z, \Theta)$. This step is called *modelling* and it is essential in data analysis. In particular, we must provide the following distributions:

- **The prior distribution $\pi(\Theta)$:** This is the distribution of our parameters $\Theta$ before we make a measurement. In general, we cannot say much about this prior distribution, except maybe some basic restriction (like for example that it can only take positive values, in the case of a mass). That is why we generally use weakly informative priors, which cover a broad range of values (like normal distributions with high variance [6]).
- **The likelihood function $f(Z|\Theta)$:** This is the conditional distribution of our observations given the real quantities $(Z|\Theta)$. This probability density function tells us how likely it is that we have observed $Z$, supposing that we know the exact parameters $\Theta$.
- **The posterior distribution $\pi(\Theta|Z)$:** This is the conditional distribution of the true quantities given our observations $(\Theta|Z)$. Is tells us how likely it is that the true parameters are $\Theta$, given that we have observed $Z$.

These three distributions are related through the Bayes formula:

$$\pi(\Theta|Z) = \frac{f(Z|\Theta)\pi(\Theta)}{\int f(Z|\Theta)\pi(\Theta)d\Theta} \tag{1}$$

Our aim is arriving at the distribution of $\Theta|Z$, that is to say, the distributions of the real parameters given our observations. With the Bayes formula, we only need to determine the likelihood distribution $Z|\Theta$ and the prior distribution of $\Theta$.

## 1.2 Simulation and the Hamiltonian Monte Carlo method in Stan

If $\Theta$ is a $K-$dimensional vector of random variables, a $K-$dimensional integral appears in the denominator of the Bayes formula. In most cases of interest, $K$ is a large number, so solving such integral is very computationally expensive. For example, if we have observations of a physical variable $Y$ for 50 galaxies, then $Y_{\text{obs}}[i] \in Z$ for $i = 1, ..., 50$, where $Y_{\text{obs}}[i]$ is the observation of the $i$-th galaxy. Then, for the real values of each galaxy $Y_{\text{real}}[i] \in \Theta$, so K is at least 50.

Therefore, instead of finding the expression for 1, we can use a simulation method instead. The simulation consists of generating values of $\Theta$ that follow the distribution $\Theta|Z$, despite not knowing the exact expression for such distribution. In particular, we will only need to know the value of the Bayes formula 1 up to a multiplicative constant that can depend on the data $Z$ (like the denominator). By generating these random values, we create a sample of the distribution, with which we can later find any quantity of interest, including the mean and the standard deviation.

The most common method to do so is the Metropolis-Hasting algorithm [3]. This is a Markov Chain Monte Carlo method, which means that at each iteration we move from one point $\Theta^{(1)}$ to another $\Theta^{(2)}$ with a probability based on the quotient (which does not depend on the integral):

$$\frac{\pi(\Theta^{(2)}|Z)}{\pi(\Theta^{(1)}|Z)} = \frac{f(Z|\Theta^{(2)})\pi(\Theta^{(2)})}{f(Z|\Theta^{(1)})\pi(\Theta^{(1)})} \tag{2}$$

The higher this quotient, the more likely we will move from $\Theta^{(1)}$ to $\Theta^{(2)}$. Therefore, the algorithm creates a sampling with many points in the areas of higher probability, and less points in the areas of lower probability, eventually converging to the distribution $\pi(\Theta|Z)$.

In this thesis, we will use a variant of the Metropolis-Hasting algorithm, which is the Hamiltonian Monte Carlo (HMC) algorithm [4]. This algorithm treats the function $U(\Theta) = -\log \pi(\Theta|Z)$ as a potential for the sample points $\Theta^{(i)}$. At each iteration, the point $\Theta^{(1)}$ is given a random momentum $\vec{p}$, and $\Theta^{(1)}$ moves in the direction of $\vec{p}$ a certain length, that depends on the shape of $U(\Theta)$. If $\vec{p}$ points towards a region of high probability ($\pi(\Theta|Z)$ is large), $\Theta^{(1)}$ takes a big step towards that region, with a movement that is based on hamiltonian mechanics. But if $\vec{p}$ points towards a region of low probability ($\pi(\Theta|Z)$ is small), $\Theta^{(1)}$ will take a small step in this iteration. Eventually, the sample also converges to the distribution $\pi(\Theta|Z)$. This algorithm will be implemented in Stan [5], which is a specific software to generate samplings using the HMC method, and run in R.

## 2 Objectives

We want to estimate masses of black holes based on observations of their host galaxies. Measuring black hole masses directly is observationally very expensive, since we must measure movements of objects in the center of the galaxies (our field of vision is one second of arc or even less). On top of that, sometimes we can only get an upper bound estimation of the mass. In contrast, many galaxies' variables are very easy to measure. The main objective of this thesis is to reach a method to estimate the black hole masses without having to measure it directly, simply by looking at its host galaxy's variables. Additionally, using Bayesian Statistics we also want to generate accurate probability distributions for the masses of black holes in the cases where we only know an upper bound. The key will be using our observations $Z$ to derive a posterior distribution for the true quantities $\Theta$. To get to the distribution $\Theta|Z$, we will work on the following objectives:

1. Creating and testing a model to analyse linear regressions where some observed data points are upper bounds of the real values.
2. Creating a model to analyse observed data $Z = (\mathcal{X}^{(1)}_{\text{obs}}, ..., \mathcal{X}^{(M)}_{\text{obs}}, Y_{\text{obs}})$ jointly, where for each $i = 1, ..., M$ there are linear dependencies $Y_{\text{obs}} \overset{\text{lin}}{\sim} \mathcal{X}^{(i)}$.

3. Applying the previous model to study galaxies' black-hole masses, assuming that there are linear dependencies between the masses ($Y$) and certain galaxies parameters ($\mathcal{X}^{(i)}$). Comparing the joint model to the the individual fits $Y_{\text{obs}} \overset{\text{lin}}{\sim} \mathcal{X}^{(i)}$ for only one $i$. In particular, we will compare the joint and individual regression lines, the estimation of the real values $Y_{\text{real}}$, and the difference between $Y_{\text{obs}}$ and $Y_{\text{real}}$ for each galaxy and each model (individual and joint).

# 3 Methodology I: Upper bound and joint fit models

## 3.1 The log probability function in Stan

Stan simulates values of certain parameters $\Theta$ given some observations of variables $Z$. The exact expression for $\pi(\Theta|Z)$ need not be known, it must only be calculated up to a multiplicative constant $C(Z)$ that does not depend on $\Theta$: $\pi(\Theta|Z) = C(Z)\Phi(\Theta, Z)$, where $\Phi$ may be any function (not necessarily a probability function). In the cases were it is evident, we will be flexible with the notation and use $\pi(\cdot)$ for both the joint and marginal distributions of $\Theta$, and $f(\cdot)$ for the joint and marginal distributions of $Z$. There are three important observations that we must make to derive the models for the rest of the thesis.

### 3.1.1 Independence of $Z_1$ and $\Theta$

Suppose that there exists a $k$-dimensional vector $Z_1$ and a $(n-k)$-dimensional vector $Z_2$ such that $Z = (Z_1, Z_2)$, and $Z_1$ is independent of $\Theta$. Then:

$$\pi(\Theta|Z) = C(Z)f(Z_2|Z_1, \Theta)\pi(\Theta) \tag{3}$$

Indeed:

$$\pi(\Theta|Z) = \frac{f(Z_1, Z_2|\Theta)\pi(\Theta)}{f(Z)} = \frac{f(Z_2|Z_1, \Theta)f(Z_1|\Theta)}{f(Z)} = f(Z_2|Z_1, \Theta)\pi(\Theta)\frac{f(Z_1)}{f(Z)}$$

Where in the last step we have used the independence of $Z_1$ and $\Theta$. This will be used to incorporate observation errors (which are independent of $\Theta$) to our Stan model's target formula.

### 3.1.2 Conditional independence of $Z_1$ and $Z_2$ given $\Theta$

Suppose now that $Z = (Z_1, Z_2)$, and that $Z_1$ is conditionally independent of $Z_2$ given $\Theta$. Then:

$$\pi(\Theta|Z) = C(Z)f(Z_1|\Theta)f(Z_2|\Theta)\pi(\Theta) \tag{4}$$

Indeed:

$$\pi(\Theta|Z) = \frac{f(Z_1, Z_2|\Theta)\pi(\Theta)}{f(Z)} = \frac{f(Z_1|\Theta)f(Z_2|\Theta)\pi(\Theta)}{f(Z)}$$

Where in the last step we used the conditional independence of $Z_1$ and $Z_2$ given $\Theta$. It is important to note that $Z_1$ and $Z_2$ need not be independent, they only need to be *conditionally* independent given $\Theta$. For example, if we are modelling a linear regression $Y \overset{\text{lin}}{\sim} X$ with observational errors in $X$ and $Y$, our data will consist of the observations pairs $Z_1 = (X_{\text{obs}}, X_{err})$ and $Z_2 = (Y_{\text{obs}}, Y_{err})$. In this case, $X_{\text{obs}}$ and $Y_{\text{obs}}$ are not independent if we are looking for a linear regression $Y \overset{\text{lin}}{\sim} X$.

However, if we take as parameters the real values $X_{\text{real}}$, $Y_{\text{real}} \in \Theta$, then $X_{\text{obs}}|(X_{\text{real}}, X_{err})$ and $Y_{\text{obs}}|(Y_{\text{real}}, Y_{err})$ are independent (where we have included $X_{err}$ and $Y_{err}$ as conditioning variables with formula 3). This conditional independence comes from the fact that $X_{\text{obs}} \sim \mathcal{N}(X_{\text{real}}, X_{err})$, so once we know $X_{\text{real}}$ and $X_{err}$, $Y_{\text{obs}}$ does not add any information on the value of $X_{\text{obs}}$ (and vice versa for $Y_{\text{obs}}$).

### 3.2 Joint fit model

Now, we will develop the joint model for our analysis. We will first focus on the case where we have two separate variables $\mathcal{X}^{(1)}$ and $\mathcal{X}^{(2)}$, with which we want to infer some physical quantities $\Theta$. For example, we might be interested in finding the values for the black hole masses of some galaxies ($Y \in \Theta$), and we have two different galaxy's variables with which we want to infer them, like its color ($\mathcal{X}^{(1)}$) and dispersion of velocities ($\mathcal{X}^{(2)}$). We can describe these situations as follows:

- Our data vector is $Z = (\mathcal{X}^{(1)}, \mathcal{X}^{(2)}, Y_{\text{obs}}, Y_{err})$, where $Y_{\text{obs}}$ is the observation and $Y_{err}$ its error.
- Our parameters are $\Theta = (\Theta_0, Y_{\text{real}})$, where $Y_{\text{real}}$ are the real values of $Y$, and $\Theta_0$ contains all the regression parameters.
- We have two different conditional distributions for $Y_{\text{real}}$. On the one hand, we know there is a linear dependence $Y \overset{\text{lin}}{\sim} \mathcal{X}^{(1)}$, so $Y_{\text{real}}|(\mathcal{X}^{(1)}, \Theta_0) \sim \mathcal{N}(\beta_0^{(1)} + \beta_1^{(1)}\mathcal{X}^{(1)}, \sigma^{(1)})$. But on the other hand, we have another linear dependence $Y \overset{\text{lin}}{\sim} \mathcal{X}^{(2)}$, so $Y_{\text{real}}|(\mathcal{X}^{(2)}, \Theta_0) \sim \mathcal{N}(\beta_0^{(2)} + \beta_1^{(2)}\mathcal{X}^{(2)}, \sigma^{(2)})$. Thus, $\Theta_0 = (\beta_0^{(1)}, \beta_1^{(1)}, \beta_0^{(2)}, \beta_1^{(2)}, \sigma^{(1)}, \sigma^{(2)})$.

In such situations, we can generate a Stan fit for each variable separately, $Y \overset{\text{lin}}{\sim} \mathcal{X}^{(1)}$ and $Y \overset{\text{lin}}{\sim} \mathcal{X}^{(2)}$. This is what we will call **individual fits**. But we can also try to make a **joint fit**, using the information of the dependence of $Y \overset{\text{lin}}{\sim} \mathcal{X}^{(1)}$ and $Y \overset{\text{lin}}{\sim} \mathcal{X}^{(2)}$ at the same time. The key of this thesis is modelling the joint distribution by introducing the following probability density function:

$$\tilde{\pi}_Z(\Theta) = \frac{\sqrt{\pi(\Theta|\mathcal{X}^{(1)}, Y_{\text{obs}}, Y_{err})\pi(\Theta|\mathcal{X}^{(2)}, Y_{\text{obs}}, Y_{err})}}{\int \sqrt{\pi(\Theta|\mathcal{X}^{(1)}, Y_{\text{obs}}, Y_{err})\pi(\Theta|\mathcal{X}^{(2)}, Y_{\text{obs}}, Y_{err})}d\Theta} \tag{5}$$

Where $\pi(\Theta|\mathcal{X}^{(1)}, Y_{\text{obs}}, Y_{err})$ and $\pi(\Theta|\mathcal{X}^{(2)}, Y_{\text{obs}}, Y_{err})$ are the standard bayesian posterior distributions. Modelling the joint distribution this way makes sense:

- If a value of $\Theta$ is very likely according to the first variable $\mathcal{X}^{(1)}$ ($\pi(\Theta|\mathcal{X}^{(1)}, Y_{\text{obs}}, Y_{err})$ is large), and it is also likely according to the second variable $\mathcal{X}^{(2)}$ ($\pi(\Theta|\mathcal{X}^{(2)}, Y_{\text{obs}}, Y_{err})$ is also large), then both posterior distributions are multiplied, and $\Theta$ will be likely in the posterior joint distribution.
- Conversely, if $\Theta$ is unlikely according to both individual distributions ($\pi(\Theta|\mathcal{X}^{(1)}, Y_{\text{obs}}, Y_{err})$ and $\pi(\Theta|\mathcal{X}^{(2)}, Y_{\text{obs}}, Y_{err})$ are both small), then the joint distribution will have a low probability for that $\Theta$.
- Finally, when a value of $\Theta$ is likely for one individual distribution but unlikely for the other, the final distribution will depend on the geometric mean: $\sqrt{\pi(\Theta|\mathcal{X}^{(1)}, Y_{\text{obs}}, Y_{err})\pi(\Theta|\mathcal{X}^{(2)}, Y_{\text{obs}}, Y_{err})}$.

Similarly, if we have $\mathcal{X}^{(1)}, ..., \mathcal{X}^{(M)}$ different variables and $Z = (Y_{\text{obs}}, Y_{err}, \mathcal{X}^{(1)}, ...\mathcal{X}^{(M)})$, then:

$$\tilde{\pi}_Z(\Theta) = \frac{\left(\prod_{i=1}^{M} \pi(\Theta|\mathcal{X}^{(i)}, Y_{\text{obs}}, Y_{err})\right)^{\frac{1}{M}}}{\int \left(\prod_{i=1}^{M} \pi(\Theta|\mathcal{X}^{(i)}, Y_{\text{obs}}, Y_{err})\right)^{\frac{1}{M}} d\Theta} \tag{6}$$

Where again $\pi(\Theta|\mathcal{X}^{(i)}, Y_{\text{obs}}, Y_{err})$ are the standard bayesian posterior distributions. Using our model hypotheses and some statistical manipulations, the target formula with which Stan generates the random sampling is:

$$\text{target}(\Theta) = \log f(Y_{\text{obs}}|Y_{\text{real}}, Y_{err}) + \frac{1}{M}\sum_{i=1}^{M} \log \pi(Y_{\text{real}}|\mathcal{X}^{(i)}, \Theta_0) + \log \pi(\Theta_0) \tag{7}$$

A derivation of this formula can be seen in the appendix 6.1.2. For come observations of formula 6, see 6.2.

## 3.3 Upper bound model

In this thesis, we will work with pairs of data points $(X_{\text{obs}}, Y_{\text{obs}})$. The subscript *obs* indicates that this is a measurement, rather than the actual true value, which we will denote with the subscript *real*. However, not all our measurements for the galaxies black holes $Y_{\text{obs}}$ portray the same information. There are two types of measurements:

- **Type 1 observations**: They are measurements that try to estimate the actual black hole mass ($Y_{\text{obs1}} \sim Y_{\text{real}}$).
- **Type 2 observations or upper bounds**: They are upper bounds for the black hole mass: ($Y_{\text{obs2}} \geq Y_{\text{real}}$).

Therefore, we need to create a model to incorporate these Type 2 observations into our Bayesian Analysis.

### 3.3.1 Stasticial model

Suppose we want to model a linear regression model with real dispersion and errors in $Y_{\text{obs}}$ but not in $X$. Also suppose that our vector $Y$ has Type 1 and Type 2 observations (direct measurements and upper bounds), and we have an input vector $T$ that informs of the type of observation:

- $T[i] = 1$ if the i-th observation of $Y$ ($Y_{\text{obs}}[i]$) is Type 1.
- $T[i] = 2$ if the i-th observation is Type 2 (an upper bound).

Therefore, our input data is $Z = (X, Y_{\text{obs}}, Y_{err}, T)$. We are interested in the parameters $Y_{\text{real}}$ (the real value of the black hole masses) and the linear regression parameters. In this case, $\Theta = (Y_{\text{real}}, \beta_0, \beta_1, \sigma)$, and we use the following distributions to model our problem:

- $Y_{\text{real}}|(X, \beta_0, \beta_1, \sigma) \sim \mathcal{N}(\beta_0 + \beta_1 X, \sigma)$, to model the linear dependence $Y \sim X$ with real dispersion.

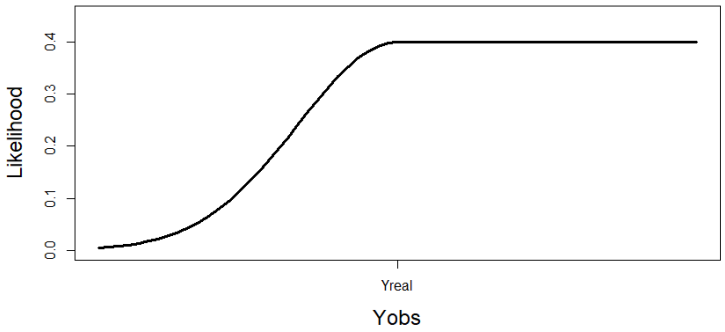



Figure 1: Likelihood function of $Y_{\text{obs}}[i]|(Y_f[i], Y_{err}[i])$ for upper bounds ($Y_{\text{err}}[i] = 1$).

- If $Y_{\text{obs}}[i]$ is Type 1, then $Y_{\text{obs}}[i]|(Y_{\text{real}}[i], Y_{err}[i]) \sim \mathcal{N}(Y_{\text{real}}[i], Y_{err}[i])$, which simply means that the observations behave like a normal distribution centered around the real value.
- If $Y_{\text{obs}}[i]$ is Type 2, we define $Y_f[i] = \max(Y_{\text{obs}}[i], Y_{\text{real}}[i])$. The key step is modelling:

$$Y_{\text{obs}}[i]|(Y_f[i], Y_{err}[i]) \sim \mathcal{N}(Y_f[i], Y_{err}[i]) \tag{8}$$

  We distinguish two cases to explain this modelling. We denote the likelihood function of $Y_{\text{obs}}[i]|(Y_f[i], Y_{err}[i])$ as $\mathcal{L}(y) = \phi_{(Y_f[i](y), Y_{\text{err}}[i])}(y)$, where $y$ is a possible value of $Y_{\text{obs}}[i]$ and $\phi_{(\mu,\sigma)}$ is the probability density function of $\mathcal{N}(\mu, \sigma)$. Then:

  - If $y \geq Y_{\text{real}}[i]$, then $Y_f[i] = Y_{\text{obs}}[i]$. In this case, $Y_{\text{obs}}[i]$ is a compatible observation with $Y_{\text{real}}[i]$ (the upper bound is above the real value) and the likelihood is $\mathcal{L}(y) = \phi_{(y,\sigma)}(y) = \frac{1}{\sqrt{2\pi} Y_{\text{err}}[i]}$. In other words, the likelihood is as high as possible (the distribution follows a normal centered on itself), and it takes the same value for all $y \geq Y_{\text{real}}[i]$ (no matter how high the upper bound, the likelihood is the same, as we are only interested in having the real value below it).
  - If $y < Y_{\text{real}}[i]$, then the observation is not compatible with $Y_{\text{real}}[i]$ (the upper bound is below the real value), therefore the likelihood decreases. The higher the separation between $Y_{\text{obs}}[i]$ and $Y_{\text{real}}[i]$, the lower the likelihood, following a normal with variance equal to the observation error of the upper bound. Such likelihood function would be $\mathcal{L}(y) = \phi_{(Y_{\text{real}}[i], Y_{\text{err}}[i])}(y)$.

  The distribution for $Y_{\text{obs}}[i]|(Y_f[i], Y_{\text{real}}[i])$ can be seen in 1.

A Stan code that implements this model is shown in 10a.

### 3.3.2 The model applied to a simple linear regression

We can test our method with a simple toy model. Suppose we have a linear dependence $Y \overset{\text{lin}}{\sim} X$, and we have $Y_{\text{obs}}$ of Type 1 and Type 2. We have 12 data points that behave like $Y \overset{\text{lin}}{\sim} X$, which are of Type 1. Next, we generate 6 data points that behave like $Y \overset{\text{lin}}{\sim} (X + 6)$, three of which will be considered as Type 1 observations and the other three as upper bounds. Next, we generate

three Type 1 observations that behave like $Y \overset{\text{lin}}{\sim} (X - 6)$, and other three Type 2 observations with the same behaviour. The data points are plotted in 2. For this model, we run Stan with the code from 10a, with which we can find the posterior distributions of $Y_{\text{real}}, \beta_0, \beta_1$ and $\sigma$ such that $Y_{\text{real}} \sim \mathcal{N}(\beta_0 + \beta_1 X, \sigma)$. We have used the means of the posterior distributions obtained with Stan to plot the values of $Y_{\text{real}}$ and the regression line $Y \overset{\text{lin}}{\sim} X$. We have also included the regression line that we would obtain if we treated all the $Y_{\text{obs}}$ as Type 1 observations. From the figure 2, we can observe:

- When the upper bounds are above the regression line (first three red points), $Y_{\text{real}}$ falls directly to the regression line, which is a very different behaviour from the previous three Type 1 observations that are above the regression line. This makes sense because the upper bounds are compatible with the expected value from the dependence $Y \overset{\text{lin}}{\sim} X$.
- However, when the upper bounds are below the regression line (last three red points), $Y_{\text{real}}$ does not rise to the regression line, because we have information that tells us it should be below it. Indeed, it remains at lower values than the Type 1 observations that are below the line, because upper bounds are more restrictive than Type 1 observations (they tell us that the value should be *below* a point, not *around* it).
- As a consequence, there is a very different regression line that gets closer to the lower values. This asymmetric behaviour is expected with upper bound observations and it will be essential throughout this thesis.

Therefore, we can see that this is a good method to model upper bounds.

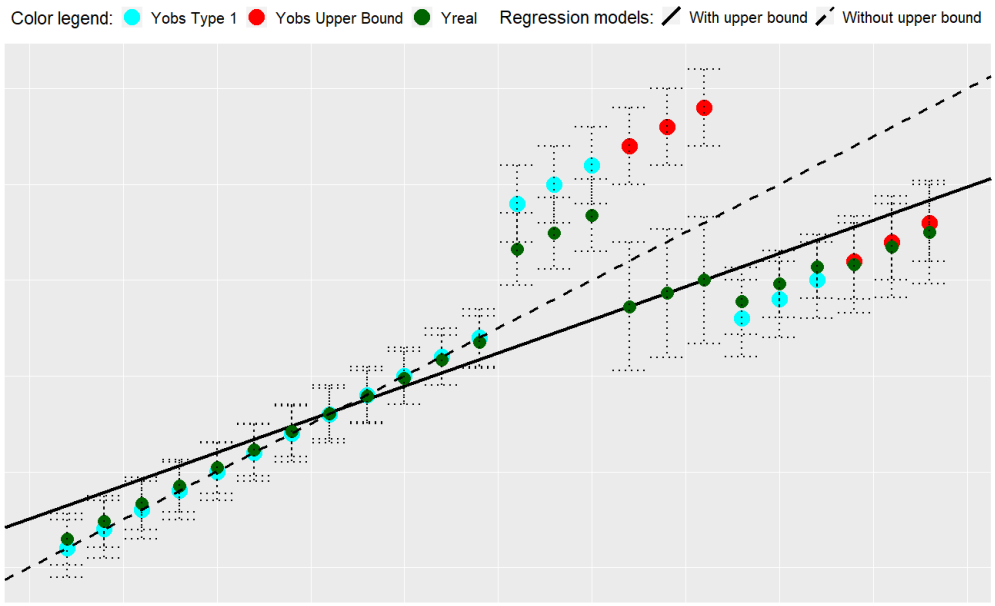



Figure 2: Upper bound toy model for a linear regression. Two regression lines are shown, the one from the upper bound model (solid) and the one that treats all data points equally (dashed).

# 4 Methodology II: Bayesian Data Analysis of Black Hole masses

## 4.1 Data

Now, we will focus on determining black hole masses, based on several physical variables of their host galaxies. We will use the data from [1] and [2], which contains a compilation of seven physical variables from 67 galaxies: the FUV-IR color (FUVcolor), the NUV-IR color (NUVcolor), the logarithm of the velocity dispersion (Logsigma), the logarithm of the effective radius (LogRe), the logarithm of the K-band luminosity (LogLk), the $3.6\mu m$-band absolute magnitude (M), and the logarithm of the mass in the K band (LogMassk). These are all variables that can be easily observed from the galaxies. For the masses of the galaxies' black holes, we use the data from [2], which provides measurements of the masses based on observation of the movements of objects in the center of the galaxies, close to the black holes. From gravitational dynamics, the masses of the black holes can be derived. Thus, our measurements of the galaxies' variables (which we will denote as $\mathcal{X}_{\text{obs}}$) are independent of the measurements of the black holes masses ($Y_{\text{obs}}$).

## 4.2 Individual fits model

We can generate individual fits for each of the galaxies' physical variables. In this case, we also have errors in X, so for the $i-$th galaxies' variable, we have a data vector $Z = (\mathcal{X}^{(i)}_{\text{obs}}, \mathcal{X}^{(i)}_{err}, Y_{\text{obs}}, Y_{err}, T)$, where $T$ indicates whether the galaxies have Type 1 or Type 2 observations for the black hole masses and $Y$ is the logarithm of the galaxy's black hole mass. Note that in this individual fit we are analyzing the $i$-th physical variable individually, not jointly with the others variables as it will be done in 4.3. On the other hand, $\Theta = (\mathcal{X}^{(i)}_{\text{real}}, Y_{\text{real}}, \Theta_0)$, where $\Theta_0 = (\beta_0, \beta_1, \sigma)$ contains the linear regression parameters such that $Y_{\text{real}}|(\mathcal{X}^{(i)}_{\text{real}}, \Theta_0) \sim \mathcal{N}(\beta_0 + \beta_1 \mathcal{X}^{(i)}_{\text{real}}, \sigma)$. To analyze this data, we must use a model that incorporates the linear regression upper bound method with errors in the $X$ variable, but this can easily be done following formulas 3 and 4 and using $\mathcal{X}^{(i)}_{\text{obs}}|(\mathcal{X}^{(i)}_{\text{real}}, \mathcal{X}^{(i)}_{err}) \sim \mathcal{N}(\mathcal{X}^{(i)}_{\text{real}}, \mathcal{X}^{(i)}_{err})$. Weakly informative priors must also be added for $\beta_0, \beta_1, \sigma$ and $\mathcal{X}^{(i)}_{\text{real}}$. The Stan code is the same as the joint fit code with $M = 1$ (see figure 10b). The results for these individual fits are shown in the left side part of figures 4 and 5.

## 4.3 Joint fit model

Following the model from 3.2, we can write a Stan code to analyze the data from all the variables together. Now $Z = (\mathcal{X}_{\text{obs}}, \mathcal{X}_{err}, Y_{\text{obs}}, Y_{err}, T)$ and $\Theta = (\mathcal{X}_{\text{real}}, Y_{\text{real}}, \Theta_0)$, where $\mathcal{X}_{\text{obs}}$ is a $(M \times N)$-dimensional matrix of random variables, such that $\mathcal{X}^{(i)}_{\text{obs}}[j]$ is the observation for the $i$-th variable in the $j$-th galaxy (same for $\mathcal{X}_{err}$ and $\mathcal{X}_{\text{real}}$).

To write a Stan code that analyses this data, we must take into account formulas 3, 4 and 7, together with the upper bound model. In this case, $Y_{\text{obs}}|(Y_{\text{real}}, Y_{err}) \sim \mathcal{N}(Y_{\text{real}}, Y_{err})$, $Y_{\text{real}}|(\mathcal{X}^{(i)}_{\text{obs}}, \Theta_0) \sim \mathcal{N}(\beta^{(i)}_0 + \beta^{(i)}_1 \mathcal{X}^{(i)}_{\text{obs}}, \sigma^{(i)})$ and $\mathcal{X}^{(i)}_{\text{obs}}|(\mathcal{X}^{(i)}_{\text{real}}, \mathcal{X}^{(i)}_{err}) \sim \mathcal{N}(\mathcal{X}^{(i)}_{\text{real}}, \mathcal{X}^{(i)}_{err})$. Applying formula 7 with these distributions, and using weakly informative priors for $\mathcal{X}^{(i)}_{\text{real}}, \beta_0, \beta_1$ and $\sigma$; we reach the model of code 10b. We will analyse the joint and individual posterior distributions with a set of graphs.

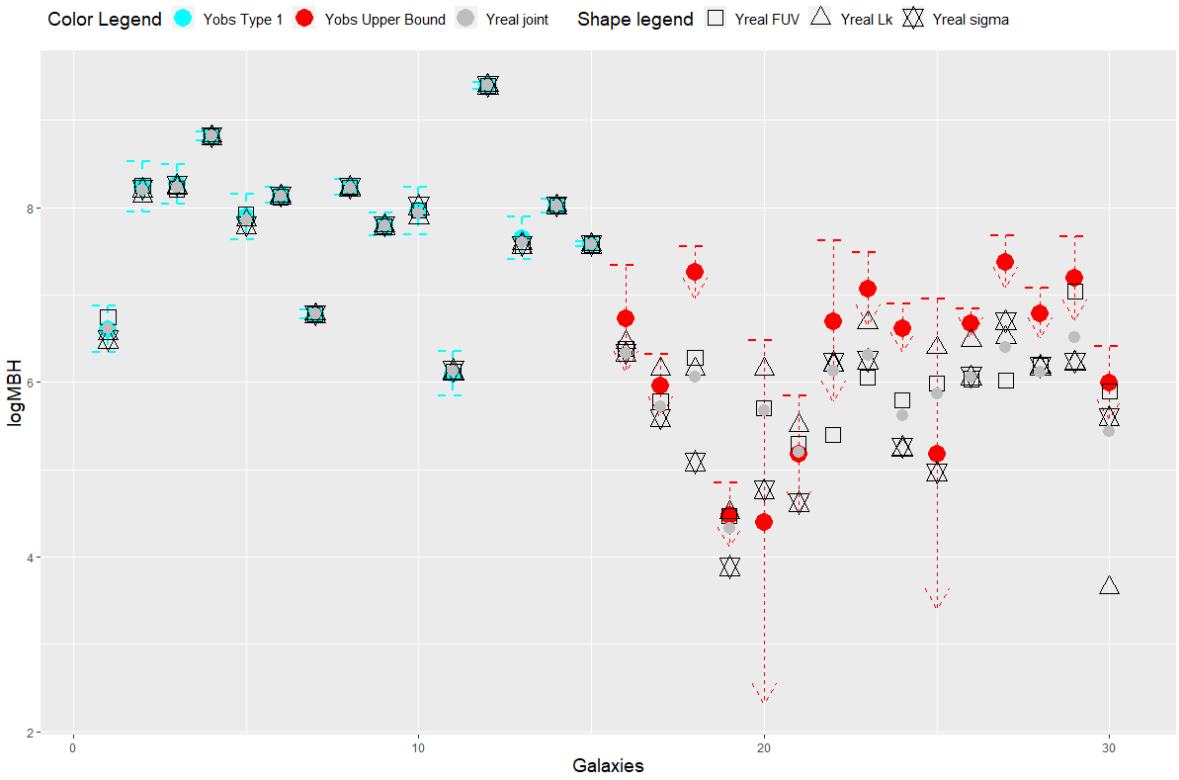



Figure 3: The values of $Y_{\text{obs}}$ and $Y_{\text{real}}$ for fifteen selected galaxies where the observation is of Type 1 and fifteen selected galaxies where an upper bound is observed. The joint distribution has only been obtained with the physical variables FUV-IR color ($Y_{\text{real}}$ FUV), K-band luminosity ($Y_{\text{real}}$ Lk) and velocity dispersion ($Y_{\text{real}}$ sigma). Each galaxy is identified with an index from one to thirty.

### 4.3.1 Graphs of individual and joint fits

In figure 3, for some galaxies we can see the difference between the inferred values of the black hole masses obtained with the individual fit methods and with the joint fit model. Also, we can compare them with the observed values of the black holes. We can reach the following conclusions:

- For galaxies of Type 1, all the values are very similar, and they lay inside the observation error range. In this case, our joint model does not behave very differently than the individual models. That is to say, in the cases where the individual fits produce very similar distributions, the joint fit follows that same distribution.
- However, in the case of the Type 2 black holes, there are significant differences between the individual fits. That is to say, the result depends on the physical variable with which we perform the individual linear fit. In this case, the joint distribution is very helpful because it takes into account the information from all the individual fits to reach a final value. Thus, the value obtained with the joint model will be more robust and less biased. We are not obtaining it from one assumption $Y \overset{\text{lin}}{\sim} \mathcal{X}^{(i)}$ for some i, rather, we are considering $Y \overset{\text{lin}}{\sim} \mathcal{X}^{(i)}$ for all i, which is why we reach an intermediate value. This means that for values of $Y_{\text{real}}$ to be valid in the joint model, we need sufficient likelihood of that value from all the individual

fits (or most of them).

- Sometimes, for the Type 2 observations, the real values for the joint and individual fits are above the observed upper bound of the galaxy. This is acceptable because such observation is not a strict upper bound, it has an observation error. Therefore, our model understands that the real upper bound could be above that point. Thus, the real values may also be above it, but falling inside the error range for the upper bound.

We can also plot graphics $Y$ against $\mathcal{X}^{(i)}$. We will do that for six of the seven variables, which are shown in figures 4 and 5. In these figures, we have indicated with an arrow the position of $(\mathcal{X}^{(i)}_{\text{real}}, Y_{\text{real}})$, compared to the observed position $(\mathcal{X}^{(i)}_{\text{obs}}, Y_{\text{obs}})$. Again, we can make three observations:

- As it happened with figure 3, the real values are very close to the observations in the Type 1 case (the arrows are very short), but they may differ significantly in the Type 2 case (the arrows are much longer).

- The joint model produces more robust values for the black hole masses than the individual models. When in the individual fits a galaxy's upper bound is above the regression line, the real value can simply fall to the line (as an example see the left-most value in the dispersion of velocities (Logsigma) graph of 4). However, in the joint fit, the model takes into account all the individual fits, therefore the real values do not simply fall to the regression line. In other words: we conclude again that our joint model fit is less biased than the individual fits.

- The linear regressions $Y \overset{\text{lin}}{\sim} \mathcal{X}^{(i)}$ are different in the individual fits and in the joint fits. If we want to infer a value of the black hole mass from a $\mathcal{X}^{(i)}$ value, the joint regression line will be less biased and therefore more reliable. This gain in robustness come at the cost of possibly underfitting the data, with regressions lines of smaller slopes.

#### 4.3.2 Graphs $Y_{\text{obs}}$ vs $Y_{\text{real}}$ and $Y_{err}$ vs $SD(Y_{\text{real}})$ in the joint model

In figure 6 we can directly observe the relationship between $Y_{\text{real}}$ of the joint model and $Y_{\text{obs}}$. Logically, the values are placed around the $Y = X$ line. For the Type 2 observations, we can expect $Y_{\text{real}} \leq Y_{\text{obs}}$, since $Y_{\text{obs}}$ is an upper bound of the observation, and this is what happens for most values. There are some values for which $Y_{\text{real}} > Y_{\text{obs}}$, but the horizontal error line always intersects the curve $Y = X$, which means that these values are compatible with the observation error. For the Type 1 observations, most values are on the line $Y = X$.

Another interesting graph is figure 7. Here, we can compare the observation error against the Standard Deviation of the posterior distribution $Y_{\text{real}}$ for the joint model. We can immediately observe that the errors for the posterior distribution are very similar to the observational errors in the Type 1 case. For the Type 2 observation, we can see that the Standard Deviation that we obtain for $Y_{\text{real}}$ is generally higher than the observation error. This makes sense because in our observations we are determining an upper bound for $Y_{\text{real}}$. Even if the error of the upper bound is very small, we still do not have much information about $Y_{\text{real}}$: we only know that it should be below one point. Thus, it makes sense that our model generates distributions with larger standard deviation than the upper bounds' observation errors.

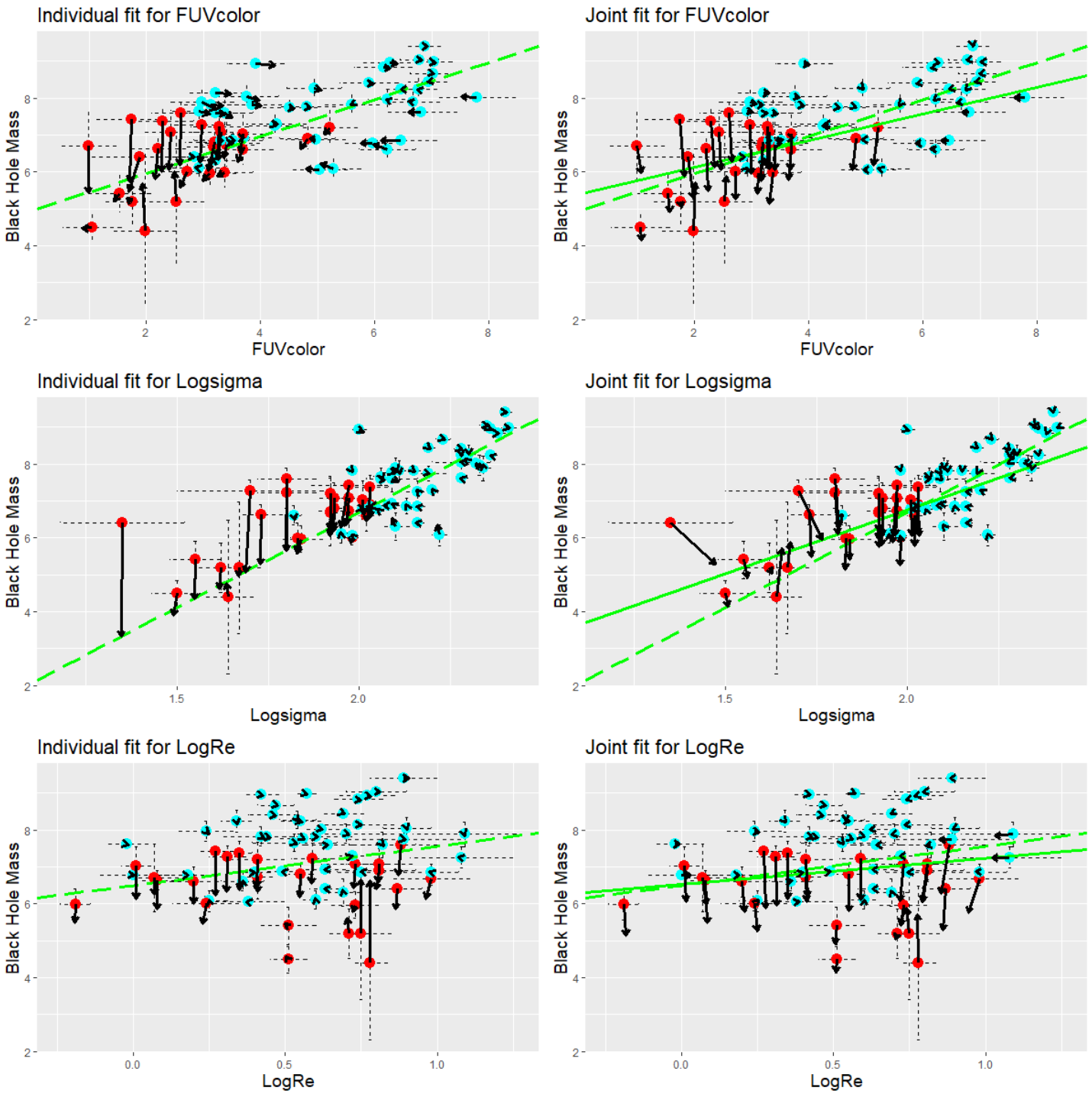



Figure 4: Individual and joint fits for three physical variables: FUV-IR color (FUVcolor), logarithm of velocity dispersion (Logsigma), and logarithm of effective radius (LogRe). The cyan dots correspond to type 1 observation and the red dots correspond to upper bound observations. Black segments start at each data point and end at the position of the $Y_{\text{real}}$ (head of the arrow). The solid green line is the joint linear regression and the dashed green line is the individual linear regression. The code for these plots can be consulted in this github repository (link).

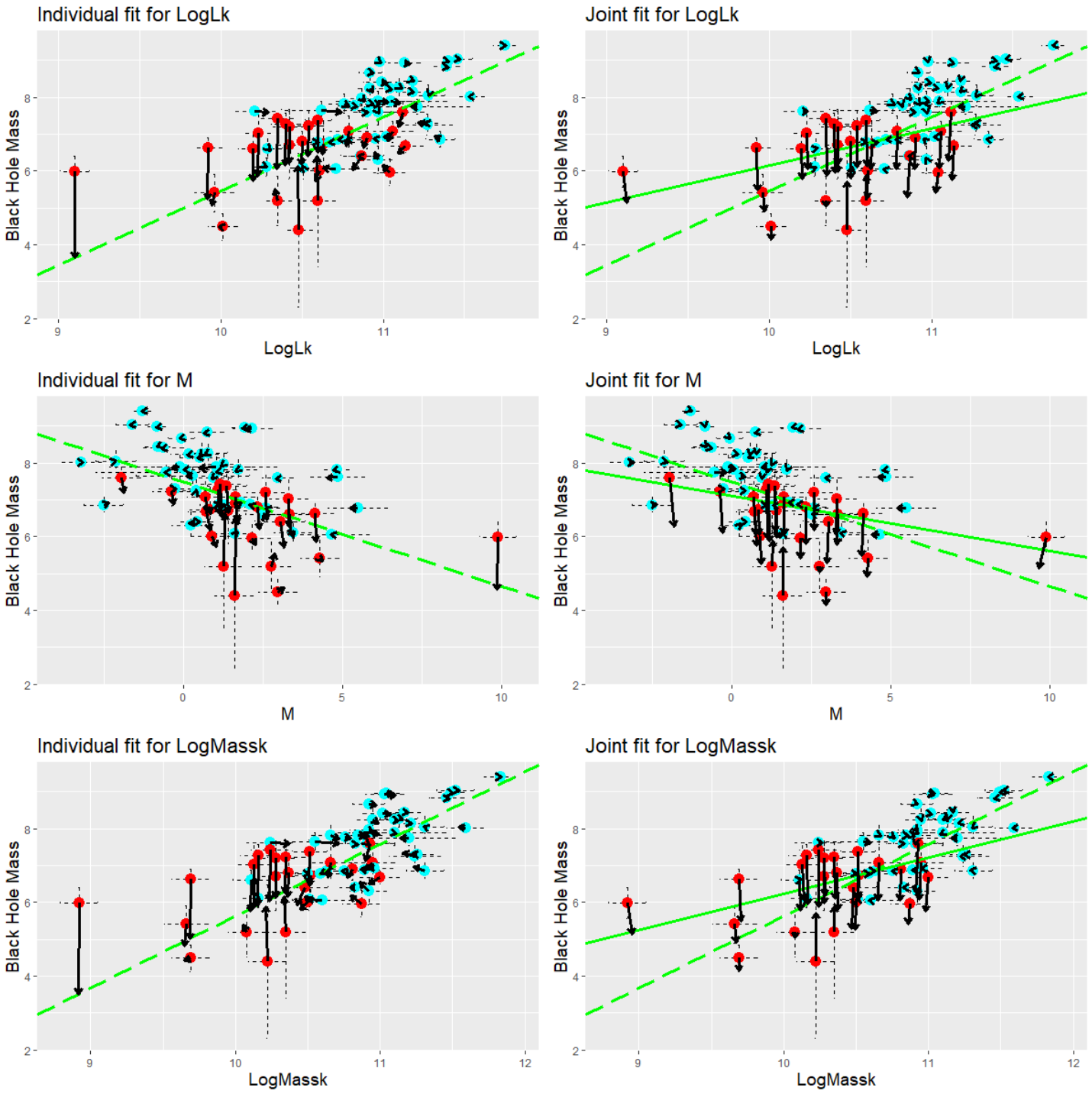



Figure 5: See Figure 4. Now, the physical variables are the logarithm of K-band luminosity(LogLk), the 3.6$\mu$ m-band absolute magnitude (M), and the logarithm of the mass in the K band (LosMassk).

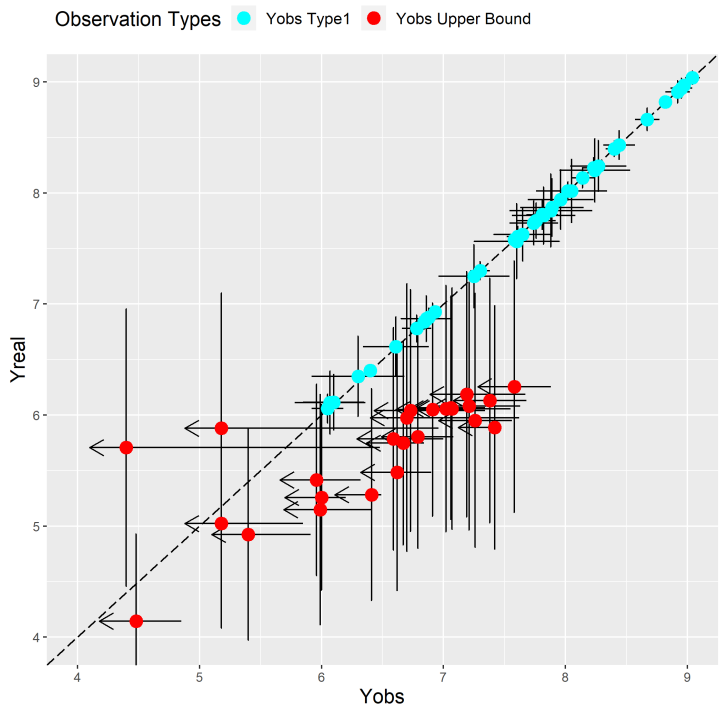



Figure 6: $E[Y_{\text{real}}]$ obtained with the joint fit model against $Y_{\text{obs}}$.

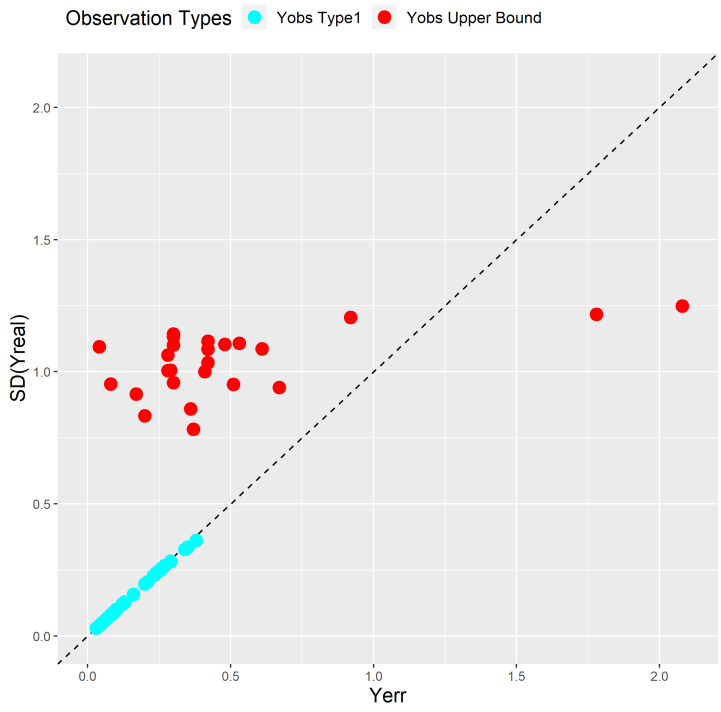
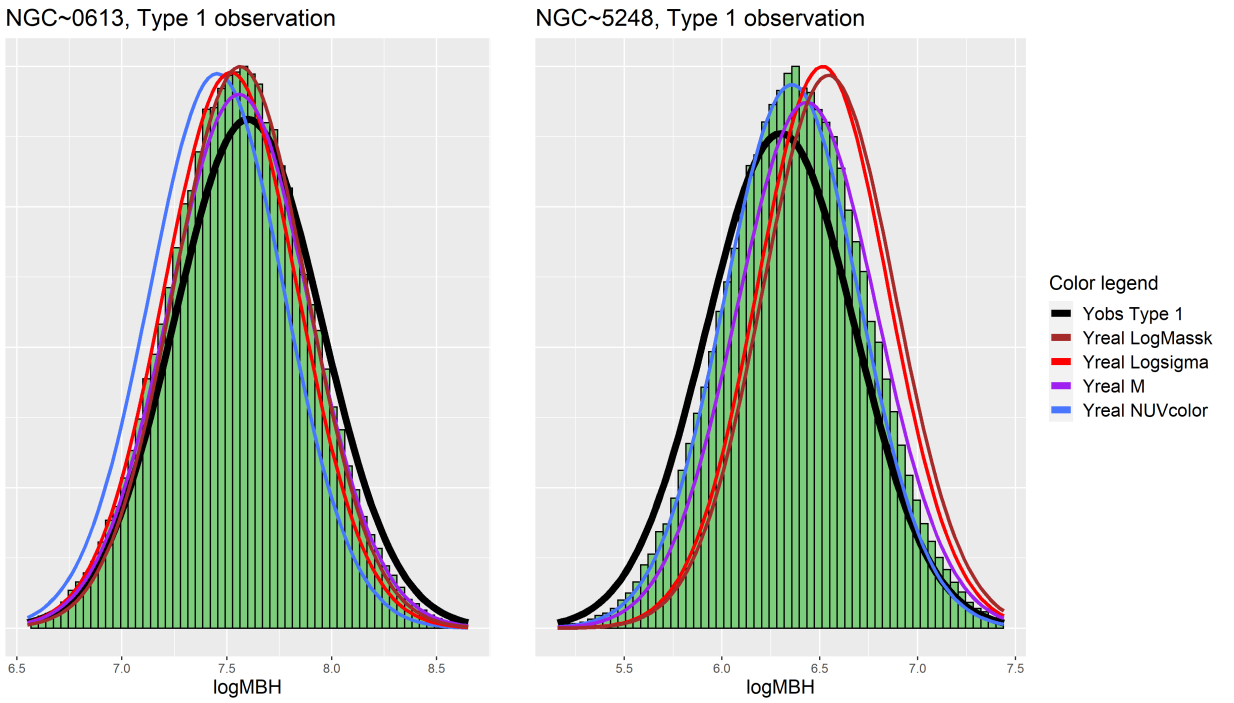



Figure 7: Comparison of observation error and Standard Deviation of the posterior distribution of $Y_{\text{real}}$ with the joint model fit.




(a) Galaxy NGC∼0613

(b) Galaxy NGC∼5248

Figure 8: Histogram of the joint model sampling of $Y_{\text{real}}$ for two Type 1 black holes, compared with the observation distribution, and four posterior distributions obtained with an individual fit of a physical variable. For these graphs, the joint model fit has only included the four variables shown in the legend. The $y$-axis scale is different for each graph.

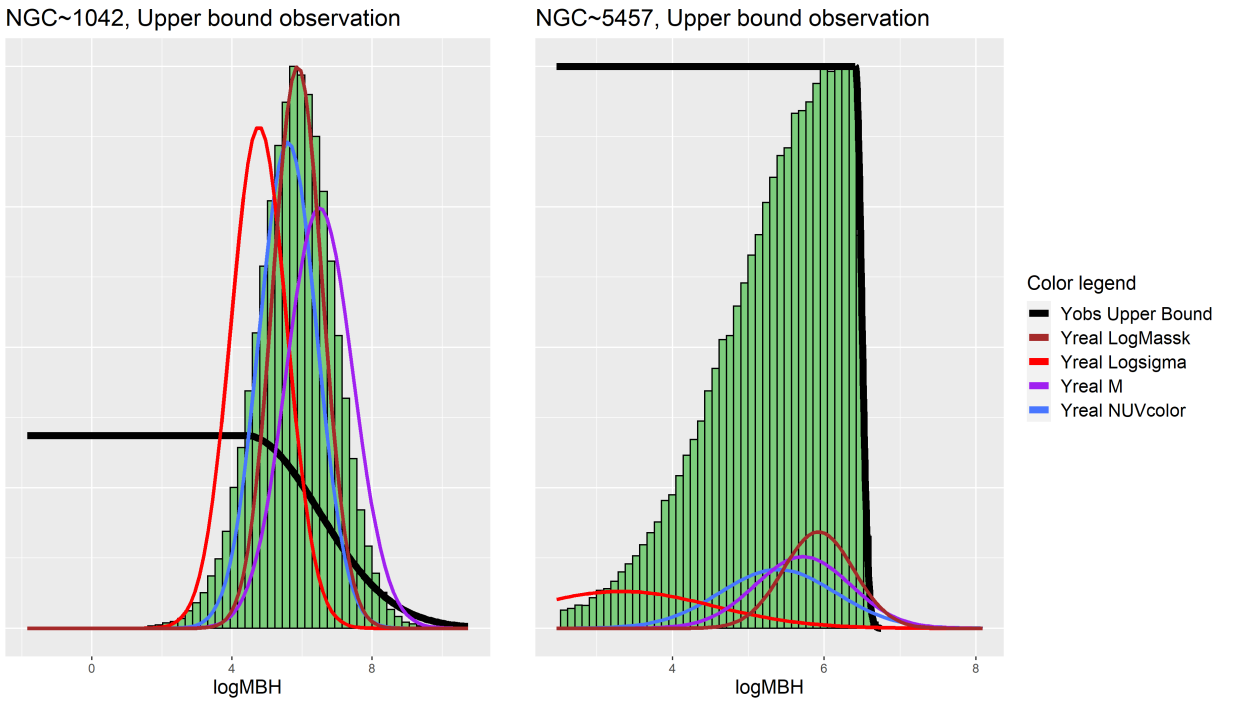
### 4.3.3 Distributions for specific galaxies

We know study the posterior distribution for some specific galaxies, and compare it to our measurements. In figure 8, we represent two histograms of the sampling of the posterior distribution of $Y_{\text{real}}$ obtained with the joint model for two Type 1 observations. With these graphs we can see how the joint model favors those $Y_{\text{real}}$ values that are likely for all individual distributions, and produces a geometric average of all of them.

In figure 9 we represent the sampling of $Y_{\text{real}}$ obtained with the joint model for an upper bound observation. In these examples we can truly see how our joint fit model works. We recall our distribution formula 6:

$$\tilde{\pi}_{\mathcal{X}_{\text{obs}}}(Y_{\text{real}}) \propto \left( \prod_{i=1}^{M} \pi(Y_{\text{real}}|\mathcal{X}_{\text{obs}}^{(i)}) \right)^{\frac{1}{K}} \tag{9}$$

This formula amplifies the probabilities of the $Y_{\text{real}}$ values that are likely for all individual fits ($\pi(Y_{\text{real}}|\mathcal{X}_{\text{obs}}^{(i)})$ is large for all $i$, thus $\tilde{\pi}_{\mathcal{X}_{\text{obs}}}(Y_{\text{real}})$ is large). However, for those values that are unlikely for some individual fits ($\pi(Y_{\text{real}}|\mathcal{X}_{\text{obs}}^{(i)})$ is low for some $i$), the probability of such values is decreased in the joint fit ($\tilde{\pi}_{\mathcal{X}_{\text{obs}}}(Y_{\text{real}})$) is small). This can be seen in the histograms of figure 9.




(a) Galaxy NGC∼1042

(b) Galaxy NGC∼5457

Figure 9: See figure 8. Here, the individual fits' distributions have been approximated with normal distributions, however, the real posterior distributions have a negative skew.

In the graph 9a, we can observe how the joint final distribution favours the $Y_{\text{real}}$ values that are likely for the four individual fits, but decreases the probability of the values that are only likely for some of them (like the ones from the dispersion of velocities' fit). The same occurs for figure 9b. This is a more extreme example where we can see that one individual fit produces an outlier value (again the dispersion of velocities' fit). The joint fit ignores these extreme values and generates a distribution around the values of $Y_{\text{real}}$ that are most likely for the three other individual fits.

# 5 Conclusions

The main result of this thesis is the generation of probability distribution for the masses of black holes for which we previously only had an upper bound estimation. The values for such masses can be consulted on Tables 1 and 2. One important observation is that the posterior distributions of these black holes' masses (Type 2 observations) are not normal, rather, they have a negative skew (the left tails are longer). This is a behaviour that appears in most of the 25 Type 2 black holes' histograms, that can be consulted on this github repository (link). Samplings of 15000 points for the joint posterior distributions (with all seven variables) for each of the 67 black holes are also available at (link).

Furthermore, the distribution we have obtained for these black holes masses is very robust. This is an advantage of the joint fit model compared to the individual fits. We can observe such robustness

in figures 3, 8 and 9. If we only consider individual linear dependencies $Y \overset{\text{lin}}{\sim} \mathcal{X}^{(i)}$ for one $i$, we are prone to biases. Outlier values heavily modify the posterior distribution, because we are assuming that $Y$ can be fully explained through a simple linear regression with $\mathcal{X}^{(i)}$. But in Astrophysics this is seldom the case. The formula for our joint model 6 amplifies the values that are most likely for all distributions, and ignores the outlier values. Because of that, we get a less biased estimation of the black holes masses, and all information comes from agreement of most individual fits, as it can clearly be seen in figure 9. This gain in robustness is particularly important for these Type 2 black holes masses, where there are considerable discrepancies among the individual fits.

It is worth noting that, although the joint posterior distribution is more robust, it generally has a greater variance than the individual posterior distributions. We obtain a more spread distribution that is less biased, and therefore less likely to be wrong. In particular, we can observe how the widely used linear relationship between the dispersion of velocities and the black hole masses might be biased when determining the masses of black holes for which there is an upper bound observation (see the red curve in figure 9). Our joint fit method utilizes the important information that the dispersion of velocities provide, but is less biased towards it in cases where there are discrepancies with the predictions from the rest of the galaxies' variables. If it is true that the black hole masses are heavily correlated with the dispersion of velocities of its host galaxies, then such bias is justified. But more Type 1 observations in wider ranges must be obtained in order to conclude that. In the regions where such Type 1 observations are scarce, relying on the dispersion of velocities' regression might lead to biased estimations.

In conclusion, this model is effective in avoiding overfitting. The regression lines that this model provides for each of the galaxies' variables are flatter than the individual fits' regression lines (see figures 4 and 5). In the regions where there are many galaxies, both lines are very similar. But in regions where we only have some outlier points, the joint regression line is more conservative when making predictions than the individual regression line. Having these robust regression lines is very helpful in Astrophysics, because with them we can predict the masses of the black holes simply by measuring variables of their host galaxies. The latter observations are much easier to make than the former, since, as we mentioned, measuring the black hole masses directly involves measuring movements of objects in the center of the galaxies (regions with 1 second of angular diameter or less in some cases). Thus, with the data from this thesis, it is possible to make unbiased estimations of the black holes masses of many galaxies.

We have also concluded that there are important differences between the posterior distribution of Type 1 observations and Type 2 observations. The first clear difference is given with figure 7. Here, we can see how the standard deviation of the posterior distribution is roughly the same as the observational error for Type 1 data points. However, for upper bounds, the standard deviation of the posterior distribution is independent of the observational error of the upper bound. This is to be expected, as we may have a very accurate observation of the upper bound, but that does not mean that we know much about the real value itself. Thus, the real value's standard deviation may still be high. Again in figure 7, a slightly positive correlation between the observation error and the posterior's standard deviation may be observed, which is reasonable since a large uncertainty in the upper bound may lead to a large uncertainty in the real value's posterior distribution. However, there are not enough data points to confirm this behaviour.

Lastly, we have seen that for Type 1 black holes, the results of our joint fit model are very similar to

the individual fits' models (see figure 6). Type 1 black help us calibrate the fit and obtain accurate posterior distributions for the Type 2 black holes. But for the Type 1 black holes themselves, the joint model generates posterior distributions that are very similar to the observed values.

# 6 Appendix

## 6.1 Stan Codes

### 6.1.1 Code for the upper bound model

In section 3.3.1, we introduced an input vector $T$ that informed whether the i-th observation was of Type 1 ($T[i] = 1$) or whether it was Type 2 ($T[i] = 2$). However, because of how our dataset is given, we will distinguish Type 1 and Type 2 observations with a different but equivalent method. We have the observations $Y_{\text{obs}}$ (which can be Type 1 or Type 2), the errors $Y_{err}$, and we will incorporate another error vector $Y_{errB}$, such that:

- $Y_{err}[i] = Y_{errB}[i]$ if the i-th observation of $Y$ ($Y_{\text{obs}}[i]$) is Type 1 ( $\iff T[i] = 1$).
- $Y_{err}[i] \neq Y_{errB}[i]$ if the i-th observation is Type 2 ( $\iff T[i] = 2$).

In this case, the only purpose of $Y_{errB}[i]$ is to determine whether $Y_{\text{obs}}[i]$ is of Type 1 or of Type 2 by comparing it with the actual error vector $Y_{err}$. Therefore, our data vector is $Z = (X, Y_{\text{obs}}, Y_{err}, Y_{errB})$. Now, for efficiency, we define two auxiliary vector $Y_{lim}$ and $Y_f$, such that:

- $Y_{lim}[i] = 0$ if $Y_{\text{obs}}[i]$ is of Type 1 (0 is an arbitrary value, it must be chosen so that it is lower than any $Y_{\text{obs}}$).
- $Y_{lim}[i] = Y_{\text{obs}}[i]$ if $Y_{\text{obs}}[i]$ is of Type 2.
- $Y_f[i] = \max(Y_{lim}[i], Y_{\text{real}}[i])$. For observations of Type 1, $Y_f[i]$ is always equal to $Y_{\text{real}}$, but for observations of Type 2, $Y_f[i] = \max(Y_{\text{obs}}[i], Y_{\text{real}}[i])$.

Now, if we implement $Y_{\text{obs}} \sim \mathcal{N}(Y_f, Y_{err})$ for all observations, we have $Y_{\text{obs}}[i] \sim \mathcal{N}(Y_{\text{real}}[i], Y_{err}[i])$ when the i-th observation is of Type 1, and $Y_{\text{obs}}[i] \sim \mathcal{N} \sim (\max(Y_{\text{obs}}[i], Y_{\text{real}}[i]), Y_{err}[i])$ when the i-th observation is of Type 2, as we wanted (section 3.3.1). The Stan code for this model can be seen in figure 10a.

### 6.1.2 Code for the joint fit model

Since using distribution 6 is novel in the field of data analysis, we will provide a careful derivation of the model we use in the Stan code (figure 10b). In the following equalities, we will denote $C_i(Z)$ as any value that depends only on $Z$.

$$\begin{aligned} \text{target}(\Theta) &= \log \tilde{\pi}_Z(\Theta) \\ &= \tfrac{1}{M} \left( \textstyle\sum_{i=1}^{M} \log \pi(\Theta | \mathcal{X}^{(i)}, Y_{\text{obs}}, Y_{err}) \right) + C_1(Z) \end{aligned} \tag{10}$$

Let $\Theta = (Y_{\text{real}}, \Theta_0)$ with $\Theta_0 = (\beta_0, \beta_1, \sigma)$, such that for each physical variable $\mathcal{X}^{(i)}$, $Y_{\text{real}} | (\mathcal{X}^{(i)}, \Theta_0) \sim \mathcal{N}(\beta_0^{(i)} + \beta_1^{(i)} \mathcal{X}^{(i)}, \sigma^{(i)})$ ($\Theta_0$ contains three $M$-dimensional vectors of random variables, where M is

the number of physicial variables). Then, for each $i$:

$$\begin{aligned}\log \pi(\Theta|\mathcal{X}^{(i)}, Y_{\text{obs}}, Y_{err}) &= \log f(\mathcal{X}^{(i)}, Y_{\text{obs}}, Y_{err}|\Theta) + \log \pi(\Theta) + C_2(Z) \\ &\overset{(1)}{=} \log f(Y_{\text{obs}}|Y_{\text{real}}, Y_{err}, \mathcal{X}^{(i)}, \Theta_0) + \log f(\mathcal{X}^{(i)}, Y_{err}|Y_{\text{real}}, \Theta_0) + \log \pi(\Theta) + C_2(Z) \\ &\overset{(2)}{=} \log f(Y_{\text{obs}}|Y_{\text{real}}, Y_{err}) + \log f(\mathcal{X}^{(i)}, Y_{err}|Y_{\text{real}}, \Theta_0) + \log \pi(\Theta) + C_2(Z)\end{aligned} \tag{11}$$

In (1) we have used the statistical formula $f(X,Y) = f(X|Y)f(Y)$, and in (2) we have used that $Y_{\text{obs}}$ and $(\mathcal{X}^{(i)}, \Theta_0)$ are conditionally independent given $(Y_{\text{real}}, Y_{err})$. The reason behind this conditional independence is that in our model we are assuming that the measurements have normal errors ($Y_{\text{obs}} \sim \mathcal{N}(Y_{\text{real}}, Y_{err})$), so once we know $(Y_{\text{real}}, Y_{err})$, $(\mathcal{X}^{(i)}, \Theta_0)$ provide no information on the value of $Y_{\text{obs}}$. Now:

$$\begin{aligned}\log f(\mathcal{X}^{(i)}, Y_{err}|Y_{\text{real}}, \Theta_0) + \pi(\Theta) &\overset{(3)}{=} \log f(\mathcal{X}^{(i)}|Y_{\text{real}}, \Theta_0) + \log f(Y_{err}|Y_{\text{real}}, \Theta_0) + \pi(\Theta) \\ &\overset{(4)}{=} \log \pi(Y_{\text{real}}|\mathcal{X}^{(i)}, \Theta_0) - \log \pi(Y_{\text{real}}|\Theta_0) + \pi(\Theta) + C_3(Z) \\ &\overset{(5)}{=} \log \pi(Y_{\text{real}}|\mathcal{X}^{(i)}, \Theta_0) + \log \pi(\Theta_0) + C_3(Z)\end{aligned} \tag{12}$$

In (3) we have used that $\mathcal{X}^{(i)}$ and $Y_{err}$ are independent, in (4) we have applied the Bayes formula and the independence of $Y_{\text{err}}$ with $(Y_{\text{real}}, \Theta_0)$, and in (5) we have used that $\pi(\Theta) = \pi(Y_{\text{real}}|\Theta_0)\pi(\Theta_0)$. Finally, applying 11 and 12 to 10, we get 7.

This joint fit model also uses the upper bound method, but the combined implementation is straightforward given the previous section. Lastly, we also have errors in $X$, but applying the formula 3, the implementation is again very easy and we omit the details. The final code can be seen in 10b.

### 6.2 Observations on the joint probability distribution model

As we commented, the use of formula 6 is novel in the field of data analysis. To understand such formula, we will discuss the quantity: $\int \left(\prod_{i=1}^{M} \pi(\Theta|\mathcal{X}^{(i)}, Y_{\text{obs}}, Y_{err})\right)^{\frac{1}{M}} d\Theta$. This has been called the affinity of the distributions $\pi(\Theta|\mathcal{X}^{(1)}, Y_{\text{obs}}, Y_{err}), ..., \pi(\Theta|\mathcal{X}^{(M)}, Y_{\text{obs}}, Y_{err})$ [7], or the Bhattacharyya distance when $M = 2$ . It is a measure of similarity between the distributions [7], and it can also be used in evaluating the features in a two-class pattern recognition problem [8]. In this thesis, formula 6 has provided very robust and accurate results, which opens the possibility to further utilize it in Astrophysical data analysis or any other kind of data analysis.

## References


[1] Dullo, B. T., Bouquin, A. Y. K., Gil De Paz, A., Knapen, J. H., and Gorgas, J. (2020). *The (black hole mass)-(color) relations for early- and late-type galaxies: red and blue sequences.* The Astrophysical Journal, 898:83. DOI.

[2] Van den Bosch, R. C. E. (2016). *Unification of the Fundamental Plane and Super-Massive Black Holes Masses.* The Astrophysical Journal, 831:134. DOI.

[3] Hastings, W.K. (1970). *Monte Carlo Sampling Methods Using Markov Chains and Their Applications.* Biometrika. 57 (1): 97–109.

```
data {
  int<lower=0> N;
  vector[N] x;
  vector[N] yobs;
  vector[N] yerr;
  vector[N] yerrB;
}
transformed data {
  vector[N] ylim;
  for(i in 1:N){
    ylim[i] = 0;
    if (yerr[i] != yerrB[i]){
      ylim[i] = yobs[i];
    }
  }
}
parameters {
  real beta0;
  real beta1;
  real<lower=0> sigma;
  vector[N] yreal;
}
model {
  vector[N] yf;
  for(i in 1:N){
    yf[i] = fmax(ylim[i],yreal[i]);
  }
  yreal ~ normal(beta0 + beta1*x,sigma);
  yobs ~ normal(yf,yerr);
  beta0 ~ normal(0.,10.);
  beta1 ~ normal(0.,10.);
  sigma ~ cauchy(0.,10.);
}
}
```

(a) Code for the upper bound model.

```
data {
  int<lower=0> N;
  int<lower=0> M;
  vector[N] yobs;
  vector[N] yerr1;
  vector[N] yerr2;
  matrix[M,N] xobs;
  matrix[M,N] xerr;
}
transformed data{
  vector[N] ylim;
  vector[M] xcen;
  vector<lower=0.>[M] xran;
  for (i in 1:N){
    ylim[i] = 0;
    if(yerr1[i] != yerr2[i]){
      ylim[i] = yobs[i];
    }
  }
  for (j in 1:M) {
    xcen[j] = mean(xobs[j,]);
    xran[j] = 5.*sd(xobs[j,]);
  }
}
parameters {
  vector[M] beta0;
  vector[M] beta1;
  vector[N] yreal;
  vector<lower=0>[M] sigma;
  matrix[M,N] xreal;
}
model {
  vector[N] yf;
  for(i in 1:N){
    yf[i] = fmax(ylim[i], yreal[i]);
  }
  for(i in 1:M){
    xreal[i] ~ normal(xcen[i], xran[i]);
    target += normal_lpdf(xobs[i] | xreal[i], xerr[i]) / M;
    target +=
    normal_lpdf(yreal | beta0[i] + xreal[i]*beta1[i], sigma[i]) /
M;}
  yobs ~ normal(yf, yerr1);
  beta0 ~ normal(0.,10.);
  beta1 ~ normal(0.,10.);
  sigma ~ cauchy(0.,10.);
}
```

(b) Code for joint fit model.


[4] S. Brooks, A. Gelman, G. Jones, and X. Meng. (2011). *Handbook of Markov Chain Monte Carlo; Chapter 5: MCMC using Hamiltonian dynamics*. arXiv.

[5] Stan Development Team. *Stan Modeling Language Users Guide and Reference Manual*. Version 2.27. `https://mc-stan.org`.

[6] Gelman A. (2020). *Prior Choice Recommendations*. Github repository.

[7] K. Matusita (1971). *Some properties of affinity and applications*. Ann. Inst. Statist. Math., 23, pp. 137–155

[8] Ray, S. (1989). *On a theoretical property of the Bhattacharyya coefficient as a feature evaluation criterion*. Pattern Recognition Letters, 9 pp. 315–319

| Galaxy | $Y_{\text{obs}}$ | $Y_{err}$ | $E[Y_{\text{real}}]$ | $SD[Y_{\text{real}}]$ | $\tilde{\mu}_3[Y_{\text{real}}]$ | Galaxy | $Y_{\text{obs}}$ | $Y_{err}$ | $E[Y_{\text{real}}]$ | $SD[Y_{\text{real}}]$ | $\tilde{\mu}_3[Y_{\text{real}}]$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| NGC∼1042 | 4.40 | 2.08 | 5.71 | 1.25 | -0.26 | NGC∼5194 | 5.96 | 0.36 | 5.41 | 0.87 | -1.16 |
| NGC∼7418 | 5.18 | 1.78 | 5.88 | 1.23 | -0.35 | NGC∼0289 | 7.38 | 0.30 | 6.12 | 1.11 | -0.81 |
| NGC∼3310 | 6.70 | 0.92 | 5.97 | 1.21 | -0.59 | NGC∼3021 | 7.26 | 0.30 | 5.95 | 1.14 | -0.83 |
| NGC∼3423 | 5.18 | 0.67 | 5.03 | 0.95 | -0.89 | NGC∼4314 | 6.91 | 0.30 | 6.06 | 0.96 | -1.07 |
| NGC∼2964 | 6.73 | 0.61 | 6.05 | 1.07 | -0.80 | NGC∼5427 | 7.58 | 0.30 | 6.26 | 1.13 | -0.80 |
| NGC∼4800 | 7.02 | 0.53 | 6.06 | 1.11 | -0.77 | NGC∼4088 | 6.79 | 0.29 | 5.81 | 1.00 | -0.99 |
| NGC∼1493 | 5.40 | 0.51 | 4.93 | 0.95 | -1.03 | NGC∼2903 | 7.06 | 0.28 | 6.06 | 1.01 | -1.01 |
| NGC∼4245 | 7.19 | 0.48 | 6.20 | 1.10 | -0.80 | NGC∼5879 | 6.62 | 0.28 | 5.48 | 1.07 | -0.93 |
| NGC∼4212 | 5.99 | 0.42 | 5.15 | 1.04 | -0.96 | NGC∼4041 | 6.00 | 0.20 | 5.26 | 0.84 | -1.31 |
| NGC∼5347 | 7.21 | 0.42 | 6.09 | 1.12 | -0.80 | NGC∼4321 | 6.67 | 0.17 | 5.75 | 0.91 | -1.20 |
| NGC∼5921 | 7.07 | 0.42 | 6.08 | 1.08 | -0.87 | NGC∼5457 | 6.41 | 0.08 | 5.29 | 0.96 | -1.17 |
| NGC∼2685 | 6.59 | 0.41 | 5.79 | 1.00 | -0.98 | NGC∼3642 | 7.42 | 0.04 | 5.89 | 1.09 | -0.91 |
| NGC∼0428 | 4.48 | 0.37 | 4.15 | 0.78 | -1.27 | | | | | | |

Table 1: Mean, Standard Deviation and Skewness of the posterior distribution of Type 2 black holes' masses, all measured in the logarithm of sun masses, except the skewness (dimensionless). These values have been estimated using a joint fit with the seven galaxies' variables, in a Stan simulation of 40000 iterations.

| Galaxy | $Y_{\text{obs}}$ | $Y_{err}$ | $E[Y_{\text{real}}]$ | $SD[Y_{\text{real}}]$ | Galaxy | $Y_{\text{obs}}$ | $Y_{err}$ | $E[Y_{\text{real}}]$ | SD$[Y_{\text{real}}]$ |
|---|---|---|---|---|---|---|---|---|---|
| NGC∼5248 | 6.30 | 0.38 | 6.35 | 0.37 | NGC∼3608 | 8.67 | 0.10 | 8.66 | 0.10 |
| NGC∼0613 | 7.60 | 0.35 | 7.56 | 0.34 | NGC∼5055 | 8.92 | 0.10 | 8.91 | 0.10 |
| NGC∼1300 | 7.88 | 0.34 | 7.84 | 0.33 | NGC∼1097 | 8.14 | 0.09 | 8.14 | 0.09 |
| NGC∼1052 | 8.24 | 0.29 | 8.20 | 0.28 | NGC∼2787 | 7.61 | 0.09 | 7.61 | 0.09 |
| NGC∼1386 | 6.07 | 0.29 | 6.11 | 0.28 | NGC∼2974 | 8.23 | 0.09 | 8.23 | 0.09 |
| NGC∼4548 | 7.25 | 0.29 | 7.25 | 0.28 | NGC∼3115 | 8.95 | 0.09 | 8.94 | 0.09 |
| NGC∼5728 | 8.05 | 0.29 | 8.02 | 0.28 | NGC∼3368 | 6.88 | 0.08 | 6.88 | 0.08 |
| NGC∼4278 | 7.96 | 0.27 | 7.94 | 0.27 | NGC∼4151 | 7.81 | 0.08 | 7.81 | 0.08 |
| NGC∼5273 | 6.61 | 0.27 | 6.61 | 0.26 | NGC∼4501 | 7.30 | 0.08 | 7.30 | 0.08 |
| NGC∼4203 | 7.82 | 0.26 | 7.79 | 0.26 | NGC∼5018 | 8.02 | 0.08 | 8.02 | 0.08 |
| NGC∼4596 | 7.89 | 0.26 | 7.87 | 0.26 | NGC∼3414 | 8.40 | 0.07 | 8.40 | 0.07 |
| NGC∼4051 | 6.10 | 0.25 | 6.12 | 0.25 | NGC∼4371 | 6.84 | 0.07 | 6.84 | 0.07 |
| NGC∼2748 | 7.65 | 0.24 | 7.62 | 0.24 | NGC∼5846 | 9.04 | 0.06 | 9.04 | 0.06 |
| NGC∼5005 | 8.27 | 0.23 | 8.24 | 0.23 | NGC∼3079 | 6.40 | 0.05 | 6.40 | 0.05 |
| NGC∼4593 | 6.86 | 0.21 | 6.87 | 0.21 | NGC∼3489 | 6.78 | 0.05 | 6.78 | 0.05 |
| NGC∼7582 | 7.74 | 0.20 | 7.73 | 0.20 | NGC∼3627 | 6.93 | 0.05 | 6.93 | 0.05 |
| NGC∼4698 | 7.76 | 0.16 | 7.75 | 0.16 | NGC∼4374 | 8.97 | 0.05 | 8.97 | 0.05 |
| NGC∼3031 | 7.81 | 0.13 | 7.80 | 0.13 | NGC∼4594 | 8.82 | 0.05 | 8.82 | 0.05 |
| NGC∼4826 | 6.05 | 0.13 | 6.06 | 0.13 | NGC∼4388 | 6.86 | 0.04 | 6.86 | 0.04 |
| NGC∼5576 | 8.44 | 0.13 | 8.43 | 0.13 | NGC∼4472 | 9.40 | 0.04 | 9.40 | 0.04 |
| NGC∼4736 | 6.78 | 0.12 | 6.78 | 0.12 | NGC∼4258 | 7.58 | 0.03 | 7.58 | 0.03 |

Table 2: Mean and Standard Deviation for Type 1 Galaxies. See table 1. In this case, the distributions are symmetric.