

**J. Zijlstra**

**An exploration of exoplanetary transit detection  
algorithms**

**Bachelor thesis**

**June 30, 2014**

**Thesis supervisors:**

**prof. dr. R.D. Gill and S.L. van der Pas, M.Sc. M.A.**



**Mathematisch Instituut, Universiteit Leiden**



# Contents

<b>Introduction</b>	<b>2</b>
<b>1 Introduction to exoplanetary research</b>	<b>4</b>
1.1 Transit method . . . . .	4
1.1.1 Geometrical alignment . . . . .	6
1.2 The presence of noise in the light curve . . . . .	6
<b>2 Box-fitting algorithm</b>	<b>8</b>
2.1 Introduction . . . . .	8
2.2 Data characteristics . . . . .	8
2.3 Explanation of the box-fitting algorithm . . . . .	9
2.3.1 Parameter estimation . . . . .	10
2.3.2 Evaluating the fits using least squares . . . . .	12
2.4 Detection significance . . . . .	15
2.5 Signal detection analysis - open problem . . . . .	17
<b>3 A Bayesian algorithm</b>	<b>19</b>
3.1 Introduction . . . . .	19
3.2 Bayesian detection algorithm for periodic signals of unknown shape . . . . .	19
3.3 Modification to the GL algorithm for the planetary case . . . . .	20
3.4 Differences between frequentist and Bayesian inference . . . . .	21
3.5 Bayesian inference . . . . .	22
3.5.1 Nuisance parameters . . . . .	23
3.5.2 Bayesian model comparison using Bayes factors . . . . .	24
3.6 Bayes factors in the planetary Bayesian algorithm . . . . .	24
3.6.1 Evidence against the null hypothesis . . . . .	28
3.7 Likelihood functions . . . . .	29
3.7.1 Likelihoods for the constant model . . . . .	30
3.8 Prior probabilities for periodic model parameters . . . . .	31
3.9 Calculations for the global Bayes factor . . . . .	32
3.9.1 Disadvantage of Bayesian algorithm with respect to detection purposes . .	32
3.9.2 Parameter estimation . . . . .	33
3.9.3 MCMC and radial velocity . . . . .	34
<b>4 Simplified “Bayesian” detection algorithm</b>	<b>35</b>
4.1 Introduction . . . . .	35
4.2 Defining the algorithm . . . . .	35
4.2.1 Maximum-likelihood approach . . . . .	36
4.2.2 Final reducing modification . . . . .	38
4.3 Similarity with box-fitting algorithm . . . . .	38
<b>Conclusion</b>	<b>40</b>
<b>References</b>	<b>41</b>



# Introduction

An exoplanet is a planet which orbits around a star that differs from our own star, the Sun. Therefore exoplanets are outside our Solar system. Mankind has long speculated about the existence of planetary systems other than our Solar system. The discussion of whether or not the Earth was unique was already debated in the 4th century BC by Greek philosophers:

*"There cannot be more worlds than one."*

— Aristotle 384 - 322 BC, *De Caelo* I.8.

*"There are infinitely many worlds both like and unlike this world of ours."*

— Epicurus 341 - 270 BC, *D.L.* 10.45.

Until the 16th century, Aristotle's Earth-centered view was the standard view of astronomy. Giordano Bruno, an Italian philosopher, put forward the view that each star is similar to our Sun and is likewise accompanied by orbiting planets. Only recently has evidence become available to confirm the philosophical speculations of the existence of planets around other stars. In 1995 the detection of the first exoplanet around a main sequence star was announced and since then the search for exoplanets has begun. Some astronomers strongly believe in extrasolar life and the existence of other Earth-like planets, bearing in mind that there are hundreds of billions of galaxies in the universe, where each galaxy, like our Milky Way, contains hundreds of billions of stars.

Different space telescopes have been and will be launched in order to detect exoplanets similar to the Earth. Since exoplanets only reflect starlight, they are billions of times fainter than their host star. So compared to their host stars, the average exoplanet is virtually undetectable. Only a few large gaseous giants orbiting stars which are much more fainter than our Sun, are directly observed by telescopes. The other exoplanets are *indirectly* detected from the light data (flux/brightness) received from the host star using *detection methods*. There are several detection methods which have been proved successful to detect exoplanets. All of these methods make use of different phenomena caused by an exoplanet which are detectable in the light signal of their host star. The most successful methods are the *radial velocity method* and the *transit method*.

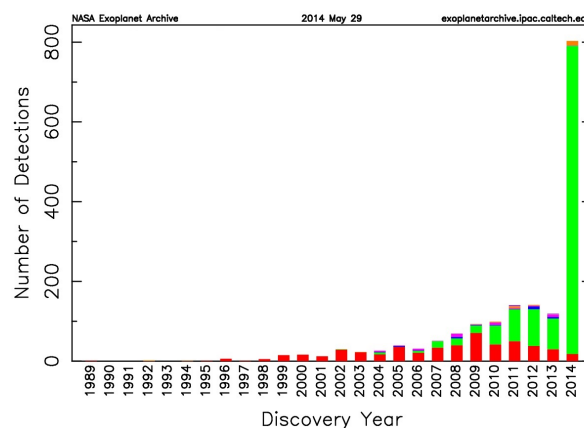


Figure 1: Number of detected exoplanets per year. The green bars represent the exoplanets detected by the transit method and the red bars represent the exoplanets detected by the radial velocity method. Figure by NASA Exoplanet Archive.

In 2014 there was a remarkable increase in the number of discovered exoplanets detected by the transit method (Figure 1), the reason being NASA's announcement on February 26; they confirmed 715 discovered exoplanets around 305 different stars by the Kepler Space Telescope, representing 40% of the total number of confirmed exoplanets (the total number of exoplanets is 1800 (The Extrasolar Planet Encyclopedia) as of June 27, 2014). Figure 1 shows us that the transit method is the most successful exoplanetary detection algorithm, therefore we will focus on the *transit method only* in this thesis. Within this transit method we will investigate different *detection algorithms* and these algorithms make use of the phenomena corresponding to the transit method. The main goal of this thesis is to explore and to compare different transit detection algorithms and we want to investigate which algorithm benefits the detection purposes most. In this thesis we will investigate algorithms based on both frequentist and Bayesian inference.

The first chapter of this thesis consists of a short introduction to exoplanetary research. We introduce the transit method and we briefly explain the characteristics and phenomena corresponding to this method. We also explain a disadvantage of this method and how this problem can be solved. This chapter serves as an introduction for certain phenomena which will be used for the detection algorithms.

In Chapter 2 we investigate the box-fitting algorithm (Kovács et al. (2002)), which is the most common approach in exoplanetary transit searches. Using trial parameters and the method of least squares, the algorithm finds the best box-shaped function approximation to the transit data. We end this chapter with an open problem containing a simulation approach to determine significant detection thresholds.

In Chapter 3 we discuss a Bayesian algorithm (Aigrain and Favata (2002)) and we give a brief introduction to Bayesian inference. The algorithm is a modification to a Bayesian algorithm for the detection of periodic modulations in X-ray data (Gregory and Loredo (1992)). We will argue that the Bayesian algorithm contains features which are an obstacle to the detection purposes.

Therefore in Chapter 4 the Bayesian algorithm has been simplified (Aigrain and Irwin (2004)), which yields to a more suitable algorithm for detection purposes based on the maximum-likelihood method. In conclusion we will argue that these simplifications and the use of maximum-likelihood estimators lead to an algorithm similar to the box-fitting algorithm because of the equivalence of the maximum-likelihood estimator and the least squares estimator.



# 1 Introduction to exoplanetary research

*“In space there are countless constellations, suns and planets; we see only the suns because they give light; the planets remain invisible, for they are small and dark.”*  
— Giordano Bruno, *Despre infinit univers si lumi* (1584).

Almost every exoplanet is not directly detectable. In order to detect these faint planets one has to exploit observable phenomena in the light data of the stars which are caused by orbiting exoplanets. In this chapter we explain the phenomenon of the *transit* in the light data used in the transit method. The algorithms investigated in this thesis are all based on this transit phenomenon. We also explain why the probability of observing a transit is very small and we briefly discuss the presence of noise which contaminates the light signal.

## 1.1 Transit method

When the exoplanet transits its host star, i.e., passes between the star and the observer (the telescope), the planet blocks a very small amount of the light signal. This results in a transit dip in the light curve of the star as illustrated in Figure 1.1. The light curve thus varies between a high out-of-transit level and a low in-transit level. Due to the orbital period of an exoplanet, a transit must be detected periodically to imply the possible presence of the exoplanet.

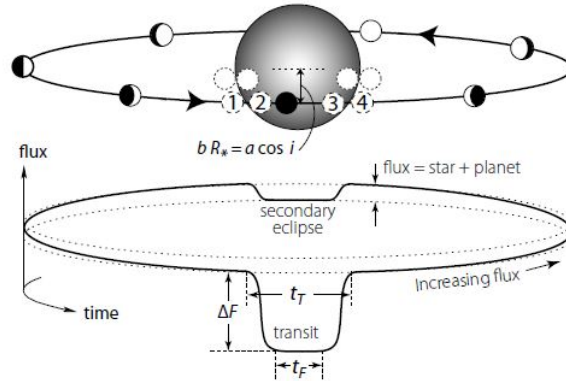


Figure 1.1: Animation of an exoplanetary transit showing a decrease in flux (brightness of the star) when the exoplanet crosses its host star. Here  $\Delta F$  is the transit depth,  $t_T$  the total transit duration and  $t_F$  the transit duration between ingress and egress (i.e., the flat part of the transit light curve). In this thesis we ignore the occurrence of the secondary eclipse. During this secondary eclipse the star is blocking the light reflected by the planet while the planet disappears behind the star. Note that also this secondary eclipse can provide important information about the planetary properties (e.g., Snellen et al. (2009)). Figure by Winn (2009).

The shape of the transit is of particular importance for estimating certain properties of the exoplanet such as the planet's radius and density. However, in this thesis we will focus on *detection* only, and therefore we are not interested in the precise shape of the transit. The interested reader is referred to Seager and Mallén-Ornelas (2003) for solutions of planet and star parameters obtained from a transit light curve.



The center of the star is brighter than the edge or limb of the star. This results in an extra drop of intensity, next to the drop caused by the planet, when the planet moves from the center of the star to its limb (Perryman (2011), p.118). This effect is known as the limb darkening effect. However, since we have little interest in the exact shape of the transit, we will ignore this effect in this thesis. Consequently, without limb darkening the transit contains a flat part at the lowest flux level. This is illustrated as the solid line in Figure 1.2.

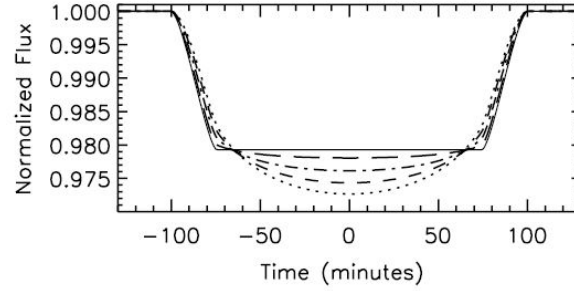


Figure 1.2: Transit light curves without limb darkening (solid line) and with limb darkening (intermittent lines for different types of stars). Figure by Seager and Mallén-Ornelas (2003).

In Figure 1.3 we illustrate some examples of transit light curves of confirmed exoplanets. The shape of the transit varies in any combination of ratios of planet and star radii. The shape of the transit also depends on an *impact parameter*, i.e., the projected distance between the planet and star centers during mid-transit. In this thesis we will not take into account this impact parameter. Again, the interested reader is referred to Seager and Mallén-Ornelas (2003).

The duration of the transit to the orbital period is relatively short; typically less than 5% (Kovács et al. (2002)). Because the fraction of the orbital period is small, the light curves in Figure 1.2 and Figure 1.3 are zoomed in. All the transit figures in this thesis represent a small section of the total light curve; they are a small percentage of the orbital period.

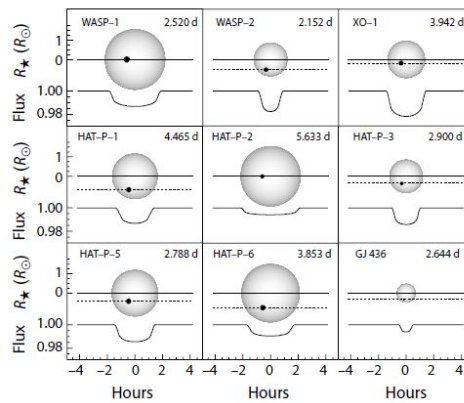


Figure 1.3: Examples of transit light curves. The planet and star sizes are shown to scale where  $R_{\odot}$  and  $R_{\star}$  stands for the planet and star radii respectively. The orbital periods are displayed in the upper right corner and the planet trajectories are shown as dotted lines. These trajectories depend on the impact parameter. Figure by Torres et al. (2008).

### 1.1.1 Geometrical alignment

A disadvantage of the transit method is that the probability of observing a transit seen from a random direction and at a random time is extremely small. The exoplanet needs to meet certain geometrical alignment criteria in order to be detectable in the light curve of its host star. Consider an exoplanet with radius  $R_p$  orbiting a star with radius  $R_\star$  at an orbital radius of  $a$ . The transit can be observed by the telescope (external observer) only if the orbital plane is sufficiently inclined with respect to the sky plane (the telescope's line of sight), see Figure 1.4.

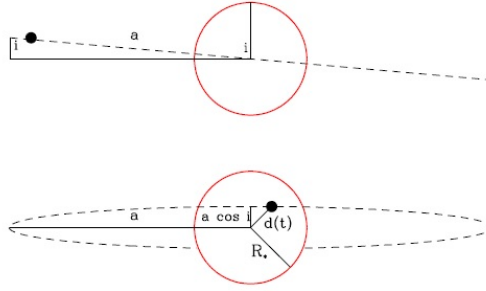


Figure 1.4: Geometry of a transit of inclination  $i$ . The dotted line is the orbital plane of the planet and the horizontal solid line is the sky plane. The top figure portrays the transit seen from the side. The bottom figure is seen from the observer's vantage point when the planet-star distance is  $d(t)$ . The exoplanet can be observed if the inclination  $i$  satisfies the geometrical criterion  $a \cos(i) \leq R_p + R_\star$ . Figure by Sackett (1999).

The probability that the inclination  $i$  satisfies the geometrical criterion for a transit is (Sackett (1999)):

$$\begin{aligned} \text{Transit probability} &= \frac{R_p + R_\star}{a} \\ &\approx \frac{R_\star}{a}. \end{aligned}$$

Only 0.5% of Earth-like planets in 1 Astronomical Unit (1 Astronomical Unit (AU) is roughly the distance from the Earth to the Sun, circa 149,6 million kilometers) orbits around Sun-like stars exhibit transits (Smith et al. (2012)). For exoplanets in orbits smaller than 1 AU the transit probability slightly increases, but overall we can say that this transit probability is very small.

In order to let the transit method be effective (looking at the small transit probability), the telescopes need to observe a very large number of stars. The Kepler telescope alone observes 100,000 stars simultaneously (Smith et al. (2012)) to enlarge the probability of observing a transit. Consequently the transit detection algorithms need to be very fast and efficient. Efficiency and speed properties of an algorithm are of great importance, bearing in mind that one is just interested in detection purposes only. In a further stage, when for instance planetary parameters need to be estimated of an already detected and confirmed exoplanet, the speed of the algorithm has lower priority.

## 1.2 The presence of noise in the light curve

The light signal we receive from an observed star is contaminated with noise. We can distinguish two different types of noise, red (correlated) noise and white (uncorrelated) noise (see Figure

1.5), which directly relates to the two different types of exoplanetary surveys: ground-based surveys and space surveys respectively. Ground-based surveys need to deal with systematics in the photometry, i.e., correlated noise in the flux measurements, caused by the Earth's atmosphere. The interested reader is referred to Pont et al. (2006) for more information about the effect of correlated noise. To detect Earth-like planets, which are of course far more interesting than giant gaseous planets, high-accuracy space-based photometry will be needed. The drop in flux caused by an Earth-like exoplanet ( $\Delta F/F = 10^{-4}$ ) is not detectable by ground-based telescopes due to the noise caused by the Earth's atmosphere according to Favata and the Eddington Science Team (2000). Therefore, in 2009 NASA launched the Kepler telescope to detect Earth-like exoplanets using a space telescope. Unfortunately two of Kepler's reaction wheels broke in 2012, therefore the telescope is unable to proceed the mission. New launches of "planet hunters" such as NASA's TESS telescope and ESA's CHEOPS telescope are planned for 2017. ESA's Gaia telescope was launched last year.

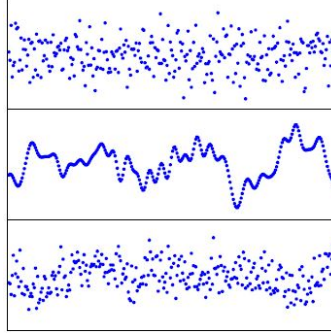


Figure 1.5: Transit light curves with white noise only (top panel), red noise only (middle panel) and white and red noise (bottom panel). Figure by Pont et al. (2006)

The algorithms applied to the light data obtained from the space surveys assume the presence of white noise only in the signal. We do not have detailed knowledge of the noise distribution, except that it has a finite variance. In Gregory (1999) it is suggested that because of the maximum entropy principle, the choice of a Gaussian distribution for the white noise would be the most conservative one, i.e., maximally noncommittal about unknown information. In Gregory (1999) the reader is referred to Jaynes (2003) for a justification of this argument. We thus assume that the white noise follows a Gaussian distribution with finite variance, which is a common assumption in the literature.

However, data obtained from the space telescopes are also contaminated by correlated noise. This red noise is caused by both instrumental and stellar variability. Sources of stellar variability can be very different, e.g., potentially large-amplitude signatures of eclipses, star spots and sub-millimagnitude changes induced by granulation. See McQuillan et al. (2012) for a detailed discussion about the effect of stellar variability on planetary light curves.

The detection algorithms examined in this thesis, are only optimized for light data embedded in white Gaussian noise. Therefore, before applying detection algorithms on planetary signals, the data set will be filtered on red correlated noise. In this thesis we will not pay further attention to the process of filtering the red noise out of the signal. We thus assume that every data set has been filtered in such a way that the assumption of white Gaussian noise is sufficient. For more information about this filtering process see McQuillan et al. (2012).

## 2 Box-fitting algorithm

### 2.1 Introduction

The first transit detection algorithm we will investigate in this thesis is the *box-fitting algorithm*. This algorithm was introduced in Kovács et al. (2002) (hereafter referred to as Paper I) and has been motivated by the increasing interest in searching for periodic transits in light curves caused by exoplanets. In Paper I, the box-fitting algorithm has proved to be more efficient for periodic transit searches than other standard period searching algorithms, available in the astronomical literature until 2002. Discrete Fourier Transformation (DST) (Deeming (1975)) and Phase Dispersion Minimization (PDM) (Stellingwerf (1978)) were until then generally accepted approaches. The box-fitting algorithm was thus an innovative algorithm in the field of exoplanetary detection. The algorithm (with and without modifications to the algorithm) has become the most common approach in transit searches ever since. In this thesis we will see why this is the case.

Other than the DST and PDM approaches, the box-fitting algorithm uses a predetermined shape of the light curve. In Section 1.1 we explained and illustrated the characteristic shape of a light curve containing transits caused by orbiting exoplanets. Due to the short duration of the transit relative to the orbital period, the signal is periodic but not sinusoidal. Sines and cosines will thus not suffice as approximations of the planetary light curve. The box-fitting algorithm therefore makes use of a step function which alternates between a “low” and a “high” state. We will see in this chapter that such a step function represents the behavior of a transit light curve, since the function is flat except during a transit dip. After we define these “transit-functions” we show how the algorithm fits the transit-functions to the data and finds the best fit using the method of least squares. In Section 2.4 we discuss the detection significance and we end this chapter with an open problem to determine significant detection thresholds using signal detection theory.

### 2.2 Data characteristics

We assume we have a data set which includes a light curve of a star with an orbiting exoplanet. For a large enough data set, we expect multiple transits in the light curve. Therefore, the light curve contains a strictly periodic signal and we will approximate the corresponding *period* by the parameter  $P$ . As described in Section 1.1, the transit signal takes on a high out-of-transit level and some lower values during the *transit phase*. Following the notation in Paper I, the level outside the transit is assumed to be constant at level  $H$ . Since we ignore limb darkening effects, we also assume the signal in the algorithm to be constant at the minimum in-transit level during the flat part (see Figure 1.2) of the transit light curve. This in-transit level is denoted by  $L$ .

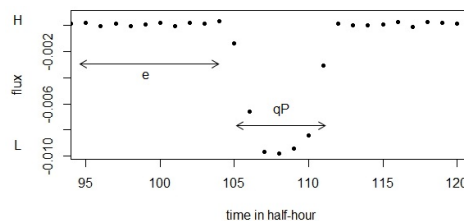


Figure 2.1: Simulated transit light curve showing the data characteristics, data simulated in **R**.

We define the time spent during the transit phase as  $qP$ , where the parameter  $q$  represents the *fractional transit length*. The transit phase is relatively short compared to the orbital period  $P$  of the exoplanet (typically less than 5%) hence we assume  $q$  to be a small number ( $\approx 0.01 - 0.05$  (Paper I)). At last the *epoch* of the transit, i.e., the position of the transit within a period, will be approximated by the parameter  $e$ .

### 2.3 Explanation of the box-fitting algorithm

The box-fitting algorithm is a transit detection algorithm developed for detection purposes only. So the main goal is to detect transits and to determine the significance of these detected transits. We already discussed to ignore the limb darkening effects in Section 1.1 since we do not want to know the exact shape of the transit. Looking at the large number of observed stars, the algorithm needs to be fast and efficient. The more accurate the approximation of the shape of the transit is, the more inefficient the algorithm becomes. We thus want to approximate the transit in a way that yields the lowest computational requirements.

The box-fitting algorithm detects planetary transits by fitting multiple transit-functions to the data. These transit-functions approximate the shape of a transit and can be thought of as the solid light curve in Figure 1.2. Because we are not interested in the shape properties of the transit, we can make a simplification to the transit-functions to make the algorithm more efficient. This simplification consists of ignoring the gradual ingress and egress phases of the transit, which carry important information about the parameters of the planetary orbit (e.g., Sackett (1999)). Fortunately, these parameters are not of interest for detection purposes. Because the lengths of the ingress and egress phases are short compared to the flat-transit phase, the omission of these phases is not expected to affect the results of the search according to Paper I. The total transit duration  $t_T$  now equals the transit duration of the flat part  $t_F$  (see Figure 1.1). This transforms the transit-function into a *box-shaped function* with two discrete values  $H$  and  $L$ , which is used by the box-fitting algorithm to fit to the transit data.

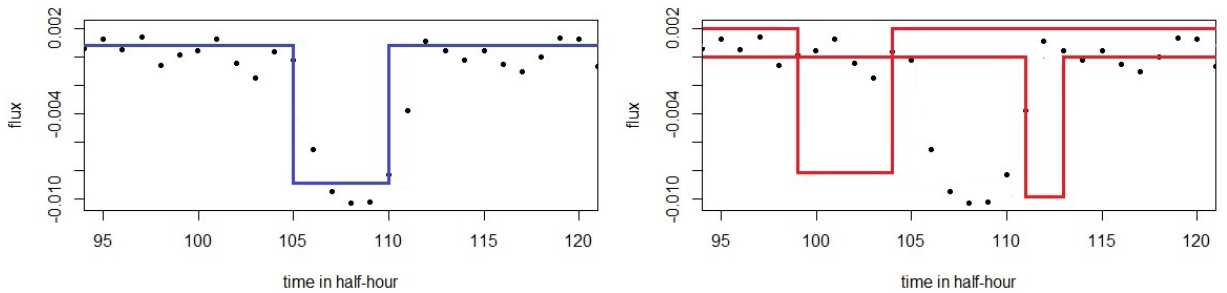


Figure 2.2: Box-shaped functions fitted to the transit data. Data simulated in **R**.

In the figures above we see different box-shaped functions fitted to the transit data, where each function contains a certain number of adjustable parameters. These adjustable parameters are the five parameters  $P$ ,  $H$ ,  $L$ ,  $q$  and  $e$ . The box-fitting algorithm aims to find the best fitting function (for instance the blue box-shaped function), i.e., we need to find the parameter set which corresponds with the best fitting box-shaped function. We can join all functions, each with different parameters, in a model. We therefore define a five-parameter model  $\mathcal{M}$  as the set of all possible model functions which are the box-shaped functions  $\mathcal{B}$ . Before we formally define the model and the model functions, we perform some modifications to the data set.

## Data centering

Before doing any kind of analysis to find an optimal parameter set for the model, we pre-process the data set by centering the data set, i.e., shift the origin of the data. We first denote the data set by the time series  $\{x_t : t \in T\}$ , where  $T = \{1, 2, \dots, n\}$  is the time interval. We can center this time series by subtracting the arithmetic average  $\mu = \frac{1}{n} \sum_{t=1}^n x_t$  from each data point  $x_t$ . This results in a centered times series with arithmetic average  $\tilde{\mu} = 0$  which is illustrated in the figure below:

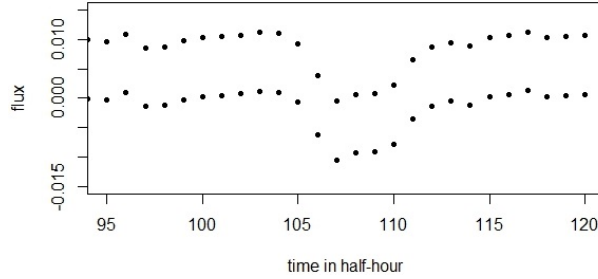


Figure 2.3: Centered transit time series (lower curve), data simulated in **R**.

This adaption causes no modifications to the structure of the time series and is only done for the sake of simplicity. For instance, since the average of the signal is zero the following equality holds:

$$(1 - q)PH + LqP = 0.$$

Later on we will see that this zero-average property results in simplifications to some derived expressions.

### 2.3.1 Parameter estimation

In the search for exoplanetary transits, one can not predict a priori the presence of exoplanetary transits in the time series  $\{x_t\}$  of a certain star, let alone the period and the other parameters of the planetary light curve if the time series contains such a light curve. We already discussed the diversity in the planetary light curves (see Figure 1.3) which corroborates our ignorance of the possible planetary parameters. We thus have to consider certain trial intervals for some parameters in the search for an optimal parameter set for the model.

### Data folding with respect to a trial period

In order to estimate the period  $P$  we perform a second modification to the time series  $\{x_t\}$ . The box-fitting algorithm makes use of the technique *data folding* to transform the time series  $\{x_t\}$  into the *folded time series*  $\{\tilde{x}_t : t \in \tilde{T}\}$  where  $\tilde{T} = \{1, \dots, \tilde{n}\}$ , which is a permutation of the original time series. The time series is folded to the period parameter  $P$ , however the real value of this parameter (if we assume the presence of a periodically transit in the time series) is of course unknown. Therefore a *trial period* is introduced in Paper I to fold the time series. Given the information of previous detected exoplanets, one can reduce the size of the interval from which one chooses the trial period. For example, in Figure 1.3 the exoplanets have orbital periods

of just a few days. These exoplanets are very close to their host star. More Earth-like planets (with larger star-planet distances) have orbits of much larger periods, typically orbital periods of several months (Aigrain and Favata (2002)).

For a given trial period  $P$  we can adapt the time index of each data point  $x_t$  by the following:

$$\tilde{t} = \text{int}(t \bmod P).$$

This results in the folded time series  $\{\tilde{x}_t\}$ . The process of data folding is illustrated in the two figures below.

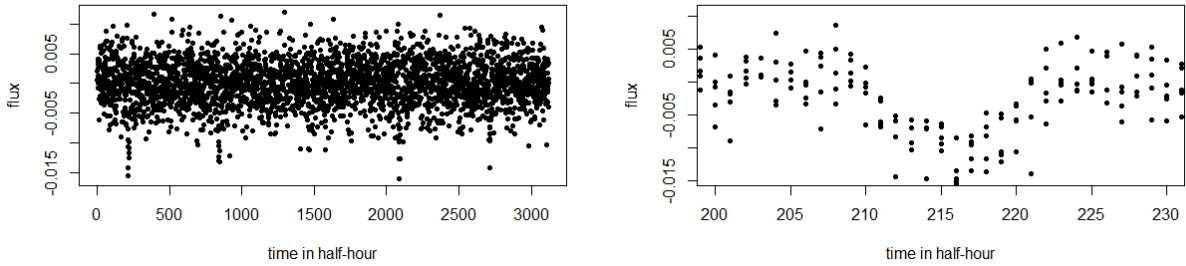


Figure 2.4: Original time series  $\{x_t\}$  (left) and the folded time series  $\{\tilde{x}_t\}$  (right). Note that the right figure is zoomed in. Data simulated in **R**.

The trial period which has been used to fold the time series in Figure 2.4, is the correct orbital period to illustrate the effect one can obtain by folding a time series with respect to the correct period. A time series folded to a trial period with a large deviation with respect to the correct period results in a very scattered folded time series. The box-shaped functions  $\mathcal{B}$  will be fitted to the folded time series  $\{\tilde{x}_t\}$  and of course less scattered folded time series yields better fits. The process of data folding thus leads to an estimate of the period  $P$ .

Note that data folding also alleviates the problem of missing data in the time series. An availability of 100%, i.e., no missing points in the time series, is not realistic according to Aigrain and Favata (2002). Gaps arise due to e.g. telemetry dropouts, spacecraft momentum dumping maneuvers, showers of solar protons during large solar flares, etc.

### Test period for epoch and fractional transit length

For a given trial period  $P$ , the algorithm fits different box-shaped functions to the folded time series  $\{\tilde{x}_t\}$ . We thus need to introduce estimates for the other four parameters in order to obtain box-shaped functions to fit to the time series  $\{\tilde{x}_t\}$ . We denote the set of the other parameters by  $\theta = (H, L, q, e)$ . Let us first define the box-shaped functions for a given trial parameter  $P$  as

$$\mathcal{B}_\theta^P(t) = \begin{cases} L & t \in [t_1, t_2], \\ H & t \in [1, t_1) \cup (t_2, \tilde{n}], \end{cases}$$

where  $[t_1, t_2]$  is the *test period* following the notation in Paper I. Varying the test period yields different values of the fractional transit parameter  $q$  and the epoch parameter  $e$ . This test period



thus provides estimates for the parameters  $q$  and  $e$ . Some restrictions for the test period are given in Paper I. The value  $t_1$  satisfies  $t_1 = 1, 2, \dots, n_{t_1}$ . The maximum fractional transit length  $q$  suspected to be present in the signal, is defined as  $\Delta t_{\max}$ . Analogously,  $\Delta t_{\min}$  is defined. The values for  $q$  are assumed to be small numbers since the transit duration to the orbital period is relatively short (less than 5%). The maximum lower index  $n_{t_1}$  consequently covers the range of  $[\tilde{n} - \Delta t_{\max}, \tilde{n} - \Delta t_{\min}]$ . Furthermore, the value  $t_2$  satisfies the inequality  $\Delta t_{\min} < t_2 - t_1 < \Delta t_{\max}$ .

### Defining the model

So for a given trial period  $P$  we obtain multiple box-shaped functions with different parameter combinations of  $P$ ,  $q$  and  $e$  by varying the test period. The algorithm searches for the best fitting function  $\mathcal{B}_\theta^P$ . In the next section we explain how the parameters  $H$  and  $L$  can be estimated from the data for given  $P$ ,  $q$  and  $e$ . Let us summarize this search in the following four-parameter model:

$$\mathcal{M}_P = \{ \mathcal{B}_\theta^P(t), \theta \in \Theta \},$$

where  $\Theta$  is the parameter space for the four-parameter combinations. The box-fitting algorithm thus contains an outer loop (trial period) and an inner loop (test period) in order to obtain different parameter combinations of  $P$ ,  $q$  and  $e$ . For each parameter  $P$ , the algorithm finds an optimal combination of the parameter set  $\theta$  for model  $\mathcal{M}_P$  which corresponds with the optimal box-shaped function  $\widehat{\mathcal{B}}_\theta^P$ . The best box-fitting function is thus the best optimal box-shaped function among the trial parameters  $P$ . We summarize this in the final model

$$\mathcal{M} = \{ \widehat{\mathcal{B}}_\theta^P(t), \text{ for all trial periods } P \}. \quad (2.1)$$

Now we understand the structure of the algorithm we will explain how the algorithm evaluates each fit. This coincides with the estimation of the parameters  $H$  and  $L$  which we can directly obtain from the data for each function by applying the *method of least squares*.

#### 2.3.2 Evaluating the fits using least squares

In Section 1.2 we argued that the light data is embedded in Gaussian noise. Each data point  $\tilde{x}_t$  thus contains a Gaussian noise term. Therefore we can decompose  $\tilde{x}_t$  in the following way:

$$\tilde{x}_t = \text{signal}_t + \epsilon_t, \quad (2.2)$$

where we assume  $\epsilon_t$  to be independent Gaussian random variables,

$$\epsilon_t \sim \mathcal{N}(0, \sigma_t^2).$$

For a given combination of parameters  $P$ ,  $q$  and  $e$  we will determine the parameters  $H$  and  $L$  directly from the folded time series by using the method of least squares. In order to apply the method of ordinary least squares, the variances of the noise components must be *homoscedastic*, i.e., each variance equals the same constant value (Rice (2007), p.554). However, this is not the case since the noise is *heteroscedastic*, i.e., the variance  $\sigma_t^2$  is not equal for every  $t$ . We thus need to assign certain weights to each data point  $\tilde{x}_t$  to obtain homoscedastic noise in order to make use of ordinary least squares. This process is also known as *weighted least squares*.



Let us define  $\tilde{\omega}_t$  as the weight assigned to data point  $\tilde{x}_t$  to obtain homoscedastic noise components, i.e.,

$$\text{Var}(\tilde{\omega}_t \epsilon_t) = c,$$

for some arbitrary constant  $c$ . The following weights are defined in Paper I:

$$\tilde{\omega}_t = \frac{\frac{1}{\sigma_t}}{\sqrt{\sum_{i=1}^n \frac{1}{\sigma_i^2}}} \quad \forall t. \quad (2.3)$$

In Paper I it is assumed that the weighted folded time series  $\{\tilde{\omega}_t \tilde{x}_t\}$  has zero arithmetic average. This assumption does not affect the results of the search and allow simple expressions for the parameters  $H$  and  $L$ . We now show that the weights defined in equation (2.3) yield homoscedastic noise:

$$\begin{aligned} \text{Var}(\tilde{\omega}_t \epsilon_t) &= \tilde{\omega}_t^2 \text{Var}(\epsilon_t) \\ &= \frac{\frac{1}{\sigma_t^2}}{\sum_{i=1}^n \frac{1}{\sigma_i^2}} \sigma_t^2 \\ &= \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i^2}}. \end{aligned}$$

Note that  $\sum_{i=1}^n \frac{1}{\sigma_i^2}$  does not depend on the index  $t$ , therefore the variances of the noise components equal the same constant. The numerator in equation (2.3) thus provides homoscedasticity and also ensures less influence on the estimation of data points with corresponding noise components with larger variability than data points with smaller variability (Rice (2007), p.593). In other words, data points with large variances weigh less. The denominator of equation (2.3) is just a normalization constant such that the equality  $\sum_{t=1}^n \omega_t^2 = 1$  holds. We will see the application of this normalization constant later on. Of course, the mean of the weighted noise components, i.e.,  $\mathbb{E}[\tilde{\omega}_t \epsilon_t]$ , is zero. So by assigning weights  $\tilde{\omega}_t$  to the data points of the folded time series the noise components become independent Gaussian random variables with zero mean and constant variance. We thus can apply ordinary least squares to the weighted folded time series to determine estimations of the parameters  $H$  and  $L$ .

For a given trial period  $P$  we have the folded time series  $\{\tilde{\omega}_t \tilde{x}_t\}$ . By varying the test period we obtain different combinations of  $q$  and  $e$  and thus different box-shaped functions. We now determine  $H$  and  $L$  of each box-shaped function with given  $P$ ,  $q$  and  $e$  using the method of least squares. For any allowed test period  $[t_1, t_2]$  we minimize the *residual sum of squares*

$$\begin{aligned} \mathcal{R}(t_1, t_2) &= \sum_{t=1}^{\tilde{n}} (\tilde{\omega}_t \tilde{x}_t - \tilde{\omega}_t \mathcal{B}_\theta^P(t))^2 \\ &= \sum_{t=1}^{\tilde{n}} \tilde{\omega}_t^2 (\tilde{x}_t - \mathcal{B}_\theta^P(t))^2 \end{aligned} \quad (2.4)$$

$$= \sum_{t=1}^{t_1-1} \tilde{\omega}_t^2 (\tilde{x}_t - H)^2 + \sum_{t=t_1}^{t_2} \tilde{\omega}_t^2 (\tilde{x}_t - L)^2 + \sum_{t=t_2+1}^{\tilde{n}} \tilde{\omega}_t^2 (\tilde{x}_t - H)^2. \quad (2.5)$$

In Paper I the relative time spent at level  $L$  is characterized by  $r = \sum_{t_1}^{t_2} \tilde{\omega}_t^2$ , i.e., by the sum of the weights of the data points at level  $L$ . Note that equation (2.4) is the *weighted sum of squared residuals* (Rice (2007), p.593). Subsequently, the estimators of  $H$  and  $L$  are the values of  $H$  and  $L$  which minimize  $\mathcal{R}(t_1, t_2)$ . In order to find these two values we take partial derivatives with respect to  $H$  and  $L$  and we set these derivatives equal to zero:

$$\frac{\partial \mathcal{R}(t_1, t_2)}{\partial H} = 2 \sum_{t=1}^{t_1-1} \tilde{\omega}_t^2 (H - \tilde{x}_t) + 2 \sum_{t=t_2+1}^{\tilde{n}} \tilde{\omega}_t^2 (H - \tilde{x}_t) = 0, \quad (2.6)$$

$$\frac{\partial \mathcal{R}(t_1, t_2)}{\partial L} = 2 \sum_{t=t_1}^{t_2} \tilde{\omega}_t^2 (L - \tilde{x}_t) = 0. \quad (2.7)$$

From equation (2.6) we obtain

$$H \left[ \sum_{t=1}^{t_1-1} \tilde{\omega}_t^2 + \sum_{t=t_2+1}^{\tilde{n}} \tilde{\omega}_t^2 \right] = \sum_{t=1}^{t_1-1} \tilde{\omega}_t^2 \tilde{x}_t + \sum_{t=t_2+1}^{\tilde{n}} \tilde{\omega}_t^2 \tilde{x}_t,$$

which implies

$$H = -\frac{s}{1-r}, \quad (2.8)$$

where

$$s = \sum_{t=t_1}^{t_2} \tilde{\omega}_t^2 \tilde{x}_t.$$

The denominator of the right hand side of equation (2.8) follows from the fact that the weights  $\tilde{\omega}_t^2$  sum up to one. The numerator of the right hand side follows from the assumption that  $\{\tilde{\omega}_t \tilde{x}_t\}$  has zero arithmetic average. From equation (2.7) we obtain

$$L \sum_{t=t_1}^{t_2} \tilde{\omega}_t^2 = \sum_{t=t_1}^{t_2} \tilde{\omega}_t^2 \tilde{x}_t,$$

and hence

$$L = \frac{s}{r}. \quad (2.9)$$

We have thus determined expressions for the last two parameters of the box-shaped function for given  $P$ ,  $q$  and  $e$ . The different trial and test periods together with the weighted least squares estimators thus give us the parameters for every possible box-shaped function. For each of these functions we can determine the sum of vertical squared deviations in equation (2.5). With the new expressions for  $H$  and  $L$  given in equations (2.8) and (2.9),  $\mathcal{R}(t_1, t_2)$  can be reduced to the expression

$$\mathcal{R}(t_1, t_2) = \sum_{t=1}^{\tilde{n}} \tilde{\omega}_t^2 \tilde{x}_t^2 - \frac{s(t_1, t_2)^2}{r(t_1, t_2)(1 - r(t_1, t_2))}. \quad (2.10)$$

For each trial period  $P$  the optimal box-shaped function  $\widehat{\mathcal{B}}_\theta^P(t)$  thus corresponds with the box-shaped function  $\mathcal{B}_\theta^P(t)$  which yields the absolute minimum of  $\mathcal{R}(t_1, t_2)$ . Note that the first term of the right hand side of equation (2.10) does not depend on the test period, and therefore one can use the second term alone to characterize the quality of the fit. Therefore the *Signal Residue* ( $SR$ ) at any given trial period  $P$  is defined in Paper I as

$$SR = \max \left\{ \left[ \frac{s(t_1, t_2)^2}{r(t_1, t_2)(1 - r(t_1, t_2))} \right]^{\frac{1}{2}} \right\}, \quad (2.11)$$

where the maximum is taken over all possible test periods. The best optimal box-shaped function, which has the optimal parameter set for the model  $\mathcal{M}$  in equation (2.1), thus corresponds with the trial period  $P$  which yields the absolute maximum of  $SR$ .

### Practical implementation

As discussed in Section 1.1.1, detection algorithms need to be very fast and efficient. In order to reduce the computational requirements, the box-fitting algorithm divides the folded time series into a certain number of bins and evaluates  $SR$  by using the binned values. The binned folded time series from Figure 2.4 is illustrated in Figure 2.5 .

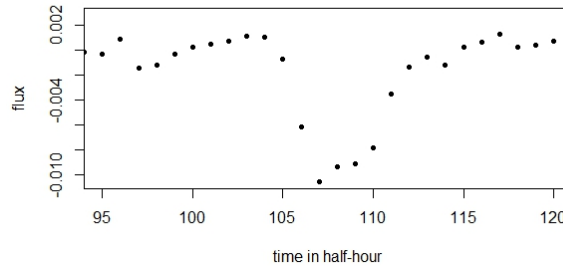


Figure 2.5: Folded time series binned, data simulated in **R**.

Note that if the bin size is too large, the transit can be fully included in the bin resulting in a decrease of detectability of the transit.

## 2.4 Detection significance

In the previous section we found the optimal box-shaped function which is the best fit for the transit data. However, we are still not able to conclude the presence of an eventual exoplanet in the data. In this section we therefore pay attention to a certain threshold which indicates the presence of an orbiting exoplanet if this threshold is exceeded.

Following the notation introduced in Paper I the *transit depth* is

$$\delta = H - L, \quad (2.12)$$

where  $H$  and  $L$  are the estimates of the parameters belonging to the optimal box function defined in equations (2.8) and (2.9). This transit depth is of course characteristic of the presence

of an exoplanet. The transit depth in the light curve caused by an orbiting planet is contaminated with Gaussian noise. These noise components thus affect the detectability of the transit depth. For simplicity reasons we assume homoscedastic noise, i.e.,  $\sigma_t^2 = \sigma^2$  for every  $t$ . This is a good approximation for space data (Aigrain and Irwin (2004)) and simplifies future necessary simulations to derive detection thresholds. So instead of focusing on the transit depth alone we need to take into account the noise presented by  $\sigma^2$ . We can derive an expression for the signal-to-noise ratio during the transit using a two-sample  $t$ -test. This ratio is defined as the *effective signal-to-noise ratio* in Paper I.

We divide our folded time series into two samples using the test period corresponding to the optimal box function. All data points within the test period belong to *sample Low* and are called the in-transit points. *Sample Low* contains  $q\tilde{n}$  data points, with  $q$  the optimal fractional transit length. The out-of-transit points together form *sample High*, which thus consists of  $(\tilde{n} - q\tilde{n})$  data points. The condition of normality for the classical two-sample  $t$ -test can be omitted since our data set is large enough. We therefore use an asymptotic  $t$ -test (Bijma et al. (2013), p.136) to investigate the presence of an exoplanet in the light curve. Using these two samples, we test the null-hypothesis (which indicates no presence of a transit in the light curve)

$$H_0 : \mu_H - \mu_L = 0$$

against the alternative hypothesis

$$H_1 : \mu_H - \mu_L > 0,$$

where  $\mu_H$  and  $\mu_L$  are the real values which are unknown to us. We use the estimates for  $H$  and  $L$  of the optimal box function as estimators for  $\mu_H$  and  $\mu_L$  respectively. Homoscedasticity was assumed so our  $t$ -statistic becomes

$$T = \frac{\hat{H} - \hat{L}}{\sqrt{\frac{\sigma^2}{\tilde{n} - q\tilde{n}} + \frac{\sigma^2}{q\tilde{n}}}}.$$

The fact that the number of data points in sample  $H$  to the number of data points in sample  $L$  are relatively large and  $q$  is small, this  $t$ -statistic can be written as follows by using equation (2.12):

$$\begin{aligned} T &\approx \frac{\delta}{\sqrt{\frac{\sigma^2}{q\tilde{n}}}} \\ &= \frac{\delta}{\sigma} \sqrt{q\tilde{n}}. \end{aligned} \tag{2.13}$$

The  $t$ -statistic defined in equation (2.13) is thus the effective signal-to-noise ratio of the transit and is a measure for the significance of the optimal fit. However, due to the fact that we have optimized certain parameters before making use of the  $t$ -test, we do not satisfy the conditions for the classical  $t$ -test. Consequently we cannot use the standard critical theoretical values of the  $t$ -distribution to reject the null hypothesis.

Therefore in Paper I, 1500 simulations are performed to determine critical values for the effective signal-to-noise ratio. To measure the detection efficiency, a Signal Detection Efficiency (SDE) was introduced, where high values of SDE correspond to efficient detections. This SDE was calculated multiple times for different effective signal-to-noise ratios (which are denoted in Paper I by  $\alpha$ ). The results of the simulations are illustrated in the Figure 2.6.

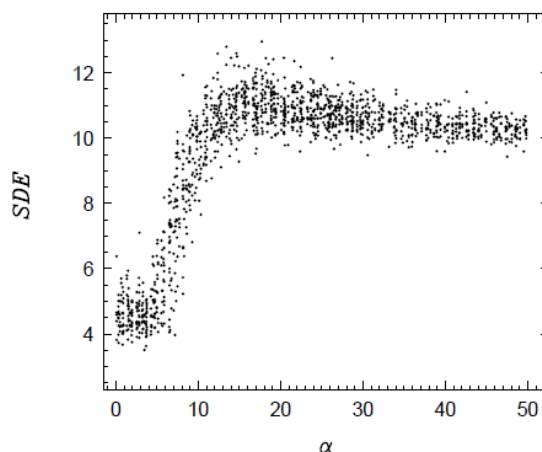


Figure 2.6: Signal detection efficiency as a function of  $\alpha$ . In the range of  $\alpha = 6 - 13$ , the SDE displays a sharp linear increase. Figure by Paper I.

When  $\alpha$  gets to the linear increase region, it turns out that the corresponding value of SDE starts to become significant according to Paper I. The final conclusion in Paper I is that in order to get a significant detection, the effective signal-to-noise ratio should *exceed* 6. According to Kovács et al. this requirement should be taken into account in future exoplanetary transit searches.

## 2.5 Signal detection analysis - open problem

In this section, due to both time and space limitations in this bachelor project, we will only briefly illustrate how one can obtain significant values for the signal-to-noise ratio using simulations *other* than the simulations described in Paper I. The simulation itself is therefore considered as an open problem.

In order to determine thresholds for the effective signal-to-noise ratio we make use of signal detection theory. We refer to Egan (1975) for more information about signal detection theory. The analysis consists of running the algorithm on two different sets of light curves. The first set consists of  $M$  simulated light curves containing only Gaussian noise and thus no transits. The other set consists of  $M$  light curves containing simulated transits. To obtain accurate results, the value of  $M$  needs to be large. However one also needs to find a proper compromise between accuracy and time constraints. For each light curve the algorithm calculates the effective signal-to-noise ratio defined in equation (2.7). The goal of this simulation is to find a certain optimal *detection threshold*. If the calculated effective signal-to-noise ratio of a light curve exceeds this threshold, this implies that the light curve contains a exoplanetary transit. When this light curve actually contains a transit this observation results in a *true positive*. A noise-only light curve exceeding the detection threshold, results in a *false positive (type I error)*: the algorithm “detects” an exoplanet when there is in fact none. On the contrary, when the algorithm does not detect an exoplanet in a light curve containing transits, this results in a *false negative (type II error)*. A *true negative* is obtained when the effective signal-to-noise ratio of a noise-only light curve is below the detection threshold.

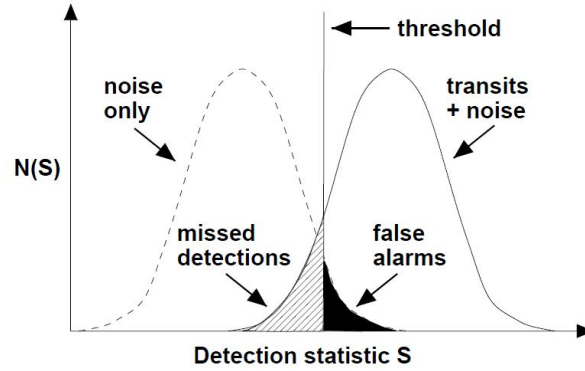


Figure 2.7: Results of a signal detection simulation displayed in a schematic diagram. Solid line: distribution of the detection statistics obtained for light curves containing both noise and transits. Dashed line: distribution of the detection statistics for light curves containing noise only. The hashed area to the left of the threshold corresponds with the type II error. The filled area to the right of the threshold corresponds with the type I error. Note that in the simulations described above the obtained distributions are discrete, resulting in a histogram. Figure by Aigrain and Favata (2002).

In order to obtain distribution histograms, the detection threshold must be chosen. The best detection threshold can be obtained from the results by a trial-and-error approach. In the simulation a certain number of detection thresholds need to be tested and subsequently the detection threshold which yields minimal type I and II errors is the best one. The location of the threshold is thus chosen as a compromise between maximizing the true positives/negatives and minimizing the false positives/negatives. Ideally the two distributions are completely separated, however in practice the two histograms show an overlap. In the simulations the critical value of 6 found in Paper I can be tested against other detection thresholds. Note that next to the distribution histograms one can also make use of Receiver Operating Characteristic (ROC) curves, which plot the true positive rate against the false positive rate. For more information about ROC curves see Fawcett (2006).

## 3 A Bayesian algorithm

### 3.1 Introduction

Statistical inference can take on many forms, but mainly it includes point estimation or interval estimation. It provides a “best guess” of an unknown parameter or it produces ranges for unknown parameters that are supported by the data (Wakefield (2013), p.27). Within statistical inference we can distinguish different inference approaches. Two commonly advocated approaches are the frequentist approach and the Bayesian approach. Since we have already investigated an algorithm based on frequentist inference (the box-fitting algorithm in Chapter 2), we explore in this chapter a planetary detection algorithm based on Bayesian inference. This algorithm is a modification of an existing Bayesian algorithm for the detection of a periodic signal of *unknown* shape, introduced in Gregory and Loredo (1992) (hereafter referred to as Paper GL). In Section 3.2 we introduce this algorithm and in Section 3.3 we look at the modifications made to this algorithm (Aigrain and Favata (2002)) such that the algorithm suits the planetary detection problem.

After the introduction of the Bayesian algorithm for the detection of exoplanets, we first give an introduction to Bayesian inference in order to explain the algorithm further using Bayesian probability theory. We end this chapter by concluding that the Bayesian algorithm is not optimal to use as a detection algorithm. However, we argue that after the detection process the Bayesian algorithm is a capable algorithm for the estimation of the planetary parameters.

### 3.2 Bayesian detection algorithm for periodic signals of unknown shape

Gregory and Loredo presented in Paper GL a new algorithm for the detection of a periodic signal in a data set when no prior knowledge of such a signal is available. This study was originally motivated to detect periodic modulations in X-ray data. The algorithm (hereafter referred to as GL algorithm) uses Bayesian probability theory to compare a constant model for the signal to a class of different periodical models (hypothesis  $H_0$  and  $H_1$  respectively). As illustrated in Figure 3.1, the periodic model is a stepwise function with  $m$  bins per period. This function describes the signal plus background noise and is capable of approximating signals of arbitrary shape. The members of the class differ from each other by the number of phase bins  $m$  which can vary between two and some upper limit and each bin has equal duration.

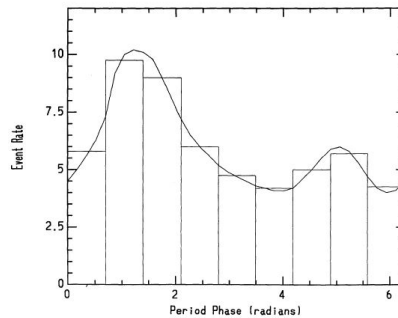


Figure 3.1: The periodic model approximates the signal with  $m$  bins per period, resulting in a stepwise function. Figure by Paper GL.

Each periodic model is denoted by  $M_m$  and has  $m + 2$  parameters:  $m$  bin levels denoted by Paper GL by  $r_j$ , a *period* parameter  $P$  and a *phase* parameter  $\phi$ . The latter specifies the location of the bin boundaries. Obviously, the constant model contains one single bin with a constant level and this results in the absence of a period and a phase.

### 3.3 Modification to the GL algorithm for the planetary case

The members of the class of periodic models in the GL algorithm are defined to approximate light curves of arbitrary shape. However, the shape of the signal in the planetary transit problem is known a priori, see Section 1.1. Therefore the GL algorithm has been modified for the planetary case in Aigrain and Favata (2002) (hereafter referred to as Paper AF). In this section we discuss this modification for the planetary case, i.e., the modification to the periodic models.

Outside the transit, we can assume that the signal is constant, again see Section 1.1. Consequently, the model can not vary outside the transit. As described in Paper AF, the modified model contains a step function with  $n + 1$  steps. The model has one “out-of-transit” bin and this bin lasts for a large fraction of any given period. The out-of-transit bin is denoted as bin 0. Bins 1 to  $n$  are the “in-transit” bins, each lasting  $d/n$  where  $d$  is defined as the *duration* of the model transit.

In Paper AF a new expression for the phase was introduced:  $\phi = 2\pi(e/P)$ . Here  $e$ , the *epoch*, is equal to the time  $t$  at the start of the first bin. However, in Paper AF it is noted that the GL method is not expected to enable phase determination directly, since there is no feature in a step function with unknown shape and equal duration steps, which is able to mark the beginning of a period. Any bin could be the first bin of a period. Therefore, in Paper AF the definition of the epoch is slightly adapted because in the transit case, one can use the transit to determine the start of a new period. The epoch  $e$  now equals the time  $t$  at the start of the first transit. The phase  $\phi$  is related to the epoch in the same way as before. Since the epoch ends in a reference point for the period and since we have information about the new duration parameter  $d$ , we can specify the location of the bin boundaries in an accurate way in the modified method. The differences between the GL algorithm and the modified algorithm for a planetary signal (hereafter referred to as AF algorithm) are illustrated in Figure 3.2.

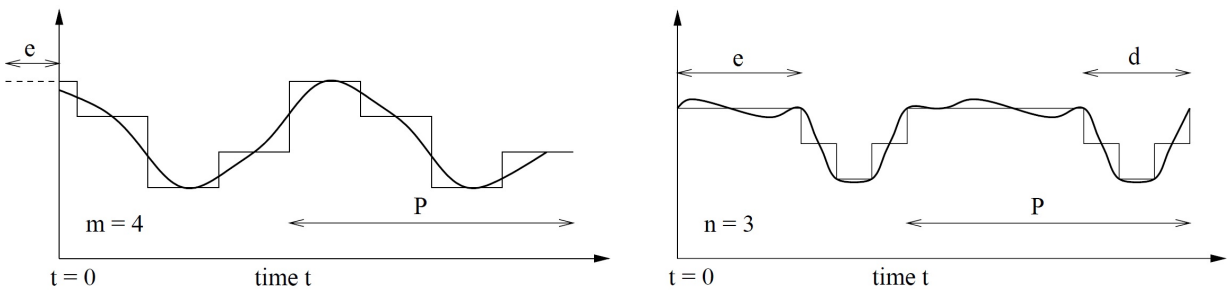


Figure 3.2: Illustration of the type of model and parameters used in the GL algorithm (left,  $m = 4$ ) and the AF algorithm (right,  $n = 3$ ). Figure by Paper AF.

The AF algorithm is the Bayesian detection algorithm we are considering in this thesis. As in the GL algorithm, the members of a class of periodic models in the AF algorithm are distinguished by the number of bins in the model. To be precise, the periodic models are distinguished by the number of bins in the transit. Outside the transit, each model has one out-of-transit bin. Each bin



corresponds to a parameter of the model. So together with the extra duration parameter  $d$ , the period  $P$  and the phase  $\phi$ , the number of parameters of a periodic model is  $n + 1 + 3 = n + 4$ . We denote each periodic model by  $M_n$  and the constant model without a transit by  $M_0$ . Because the phase  $\phi$  is related to the epoch  $e$ , one can also replace the parameter  $\phi$  by the parameter  $e$ .

### 3.4 Differences between frequentist and Bayesian inference

Before we further investigate the Bayesian algorithm, we first explain the concept of Bayesian inference and the main differences between frequentist and Bayesian inference. The frequentist point of view is based on the following postulates: (which we replicated from Wasserman (2004), p.205)

- (F1) Probability refers to limiting relative frequencies. Probabilities are objective properties of the real world.
- (F2) Parameters are fixed, (usually unknown) constants. Because they are not fluctuating, no probability statements can be made about parameters.
- (F3) Statistical procedures should be designed to have well defined long run frequency properties. For example, a 95% confidence interval should trap the true value of the parameter with limiting frequency at least 95%.

Consider the following statement:

*“The probability that this coin will land heads up is  $\frac{1}{2}$ ”.*

In the frequentist approach this statement means that if the experiment were repeated many times, the long-run average number of heads would tend to  $\frac{1}{2}$ . In the Bayesian approach, this statement is a quantification of the speaker’s uncertainty about the outcome of the experiment. The probability that the coin will land heads up may be different for different speakers. This probability can depend on the speaker’s experience and knowledge of the situation (Rice (2007), p.26). The Bayesian approach is based on the following postulates: (Wasserman (2004))

- (B1) Probability describes degree of belief, not limiting frequency.
- (B2) We can make probability statements about parameters, even though they are fixed constants.
- (B3) We make inferences about a parameter  $\theta$ , by producing a probability distribution for  $\theta$ . Inferences, such as point estimates and interval estimates may then be extracted from this distribution.

Note that there exist different philosophies within the Bayesian inference. The main schools of thoughts are the *objective* and the *subjective* Bayesian school. The latter proclaims that before the data is collected, one needs to assign a probability distribution to the parameter  $\theta$  which reflects our subjective opinion about the parameter (Wasserman (2004), p.213). Although the objective school also assigns distributions to the parameters before seeing the data, the objectivists want to minimize a person’s contribution to the inference. The a priori probability statements about parameters made by objectivists are known to be minimally informative. The motto in objective Bayesian inference therefore is: “Let the data speak for themselves”. However, the objective school also uses probability as a “degree of belief” (Berger (2006)), which can thus be confusing.

Note that the word “belief” used by objectivists does not indicate the use of their own opinions. So when we use the word “belief” hereafter, this does not imply that we speak from a subjective point of view.

While the objective school is widely accepted, the subjective school has faced a lot of criticism (Bijma et al. (2013), p.75). In this thesis we will not pay attention to discussions about this subject. Neither do we enter the debate over the conceptual superiority of Bayesian inference or frequentist inference in general. We restrict our conclusions about this superiority only to the case of the *detection* of planetary signals.

### 3.5 Bayesian inference

Let us consider some observed data set  $X$  containing  $N$  data points  $x_t$  (in this chapter we use capital  $N$  instead of  $n$  for notational reasons), which is in our case a time series describing the light curve of some observed star. In Bayesian inference we assess different competing hypotheses  $H_i$  in the light of some data  $X$ . In this thesis, these hypotheses often represent parameterized models, with parameters  $\theta_i \in \Theta$ , where each parameterized model has its own parameter space  $\Theta$ .

Different from a frequentist point of view, where parameters are treated as fixed constants, in the Bayesian approach all unknown quantities contained in a probability model for data  $X$  are treated as random variables (Wakefield (2013), p.85). Each of the competing models will be assessed by calculating the probability distribution of each model, given the data and any background information  $I$ . Also posterior distributions for the model parameters can be calculated. We will see later on that one can obtain estimates for parameters by summarizing the posterior in terms of a “best-fit” value. By looking at a simple example of a model with one parameter  $\theta$ , we show how one can derive the probability density for the model, which we will denote by  $p(\theta|X, I)$  following the notation introduced by Jeffreys (1961).

The first step in a Bayesian analysis to derive the probability distribution for the parameterized model, is to assign certain probability distributions to the model and the parameters before we obtain the data  $X$ . These distributions are called the *prior distributions* or *priors*. This prior represents our belief in the parameters of  $\Theta$  given our background information  $I$ . Note that when one uses an objective prior, which is minimally informative, the background information  $I$  has little effect on the prior.

The second ingredient needed to derive the probability density of the parameterized model, is a sampling probability for  $X$ , or a likelihood function for  $\Theta$  (Kass and Raftery (1995)). This likelihood function is the conditional distribution of  $X$  given  $\Theta$  and our background information  $I$ . For the remainder of this chapter we do not explicitly denote  $I$  in each probability density. Let us now consider the example containing the model with one parameter  $\theta$ . We denote our prior density by  $\pi(\theta)$  and we denote the likelihood function by  $p(X|\theta)$ . Before we obtain the data, we have a certain belief in  $\theta$  described by probabilities. We now want to update this belief by observing the data  $X$ . The joint density of  $X$  and  $\theta$  is

$$p(X, \theta) = p(X|\theta)\pi(\theta),$$

and the marginal density of  $X$  is

$$\begin{aligned} p(X) &= \int p(X, \theta) d\theta \\ &= \int p(X|\theta)\pi(\theta) d\theta. \end{aligned} \tag{3.1}$$

We express this updated belief in  $\theta$  in a probability distribution for  $\theta$ , in this case our only parameter, given the data  $X$ . By Bayes' theorem we obtain

$$p(\theta|X) = \frac{p(X|\theta)\pi(\theta)}{p(X)}. \quad (3.2)$$

Note that the denominator is just a normalizing constant, which ensures that the following condition is satisfied:

$$\int p(\theta|X)d\theta = 1.$$

The probability density of  $\theta$  denoted in equation (3.2) is called the *posterior density* or *posterior*. This posterior tells us how our prior belief should change after we have observed the data  $X$ . The transformation from  $\pi(\theta)$  to  $p(\theta|X)$  tells us what we have learned from the data  $X$  about  $\theta$ , so we can see the transformation itself as the evidence provided by the data (Kass and Raftery (1995)). One can see the posterior distribution as a posterior belief in  $\theta$ . It is important to note that this posterior distribution does not tell us what our beliefs should be. If we ignore the normalizing constant we can summarize the preceding result as

$$p(\theta|X) \propto p(X|\theta)\pi(\theta),$$

or in words: (Rice (2007), p.286)

$$\text{posterior distribution} \propto \text{likelihood} \times \text{prior distribution}.$$

Note that posteriors from a certain problem can be used as a prior for another problem. In Section 3.9.2 we see how one can obtain point estimates for  $\theta$  from the posterior density by making use of a *posterior mean*. One can also obtain *credible regions* from the posterior density (Paper GL), which are analogous to confidence intervals in frequentist inference.

### 3.5.1 Nuisance parameters

In the previous section we discussed an example of an one-parameter model and this resulted in a posterior density  $p(\theta|X)$ . In reality, a parameterized model can have multiple parameters and hence the posterior distribution for such a model is a joint posterior distribution for all of the parameters. Take for example a model  $M$  with parameters  $\theta$ ,  $\phi$  and  $\psi$ , then  $p(\theta, \phi, \psi|X)$  is our posterior density, which we call the *full posterior*. In some situations, one wants to focus on just a single parameter or on a subset of the parameters. The other parameters are not of interest and therefore they become nuisance parameters. So if we want to focus on the parameter  $\psi$  in our example, we need to marginalize the full posterior, i.e., we need to integrate out the nuisance parameters. This results in the posterior density for  $\psi$ ,

$$p(\psi|X) = \int \int p(\theta, \phi, \psi|X)d\phi d\theta.$$

One can thus ignore nuisance parameters very straightforward by using marginalization within Bayesian inference. Note that in equation (3.1), the normalization constant  $p(X)$  is obtained by integrating out all parameters of the joint density.

### 3.5.2 Bayesian model comparison using Bayes factors

The Bayes factor is the Bayesian alternative to the frequentist hypothesis testing and is developed by Sir Harold Jeffreys (1935, 1961). Jeffreys was concerned with the comparison of predictions made by two competing scientific hypotheses. In his work, statistical models are introduced to represent the probability of the data according to each of the hypotheses. Bayes factors are calculated using Bayes' theorem in order to make statements about which hypothesis is true. The Bayes factor is thus used in Bayesian inference to compare and test rival hypotheses. After we introduced the Bayes factors, we apply these Bayes factors to the Bayesian algorithm (AF algorithm, Section 3.3) in Section 3.6, to test the null hypothesis corresponding to the constant model  $M_0$  against the alternative hypothesis corresponding to the class of periodical models  $M_n$ .

It is often useful to consider the ratio of posterior densities corresponding to different hypotheses. In Bayesian inference this ratio is called the *posterior odds*  $\mathcal{O}_{H_i H_j}$  in favor of hypothesis  $H_i$  over hypothesis  $H_j$ . Using Bayes' theorem we can write this posterior odds as follows:

$$\mathcal{O}_{H_i H_j} = \frac{p(H_i|X)}{p(H_j|X)} \quad (3.3)$$

$$\begin{aligned} &= \frac{\pi(H_i) p(X|H_i)}{\pi(H_j) p(X|H_j)}, \\ &= \frac{\pi(H_i)}{\pi(H_j)} \mathcal{B}_{H_i H_j}, \end{aligned} \quad (3.4)$$

where we define

$$\mathcal{B}_{H_i H_j} = \frac{p(X|H_i)}{p(X|H_j)} \quad (3.5)$$

as the *Bayes factor* (Kass and Raftery (1995)). So in words, we can write

$$\text{posterior odds} = \text{prior odds} \times \text{Bayes factor}.$$

### 3.6 Bayes factors in the planetary Bayesian algorithm

In this section we apply the aspects of Bayesian probability theory introduced in the previous sections to the Bayesian algorithm. Our transit data  $X$  does or does not contain a planetary light curve. We therefore assume that our data  $X$  has arisen under one of the hypotheses  $H_0$  or  $H_1$  according to the probability densities  $p(X|H_0)$  and  $p(X|H_1)$ . We define the hypotheses as follows:

$$H_0 : \text{data } X \text{ does not contains planetary transits}$$

and the alternative hypothesis

$$H_1 : \text{data } X \text{ contains planetary transits.}$$

In the Bayesian detection algorithm the null hypothesis corresponds with the constant model  $M_0$  and the alternative hypothesis corresponds with the whole class of periodic models  $M_n$ . The posterior distribution for the null hypothesis is thus

$$p(H_0|X) = p(M_0|X) \quad (3.6)$$

and we can write the posterior distribution for the alternative hypothesis as follows (Paper GL):

$$p(H_1|X) = \sum_{j=1}^{\nu} p(M_j|X), \quad (3.7)$$

where  $\nu$  is the total number of periodic models considered. Note that the periodic models  $M_n$  are also hypotheses and these hypotheses together form hypothesis  $H_1$ . We will thus consider the posteriors of the constant and periodic models:

$$p(M_i|X) = \frac{\pi(M_i)p(X|M_i)}{p(X)}.$$

Since we want to compare our hypotheses  $H_0$  and  $H_1$ , i.e., comparing the constant model with the class of periodic models, we want to compare the models without knowledge of the parameter values. Therefore we write  $p(X|M_i)$  as follows:

$$\begin{aligned} p(X|M_i) &= \int p(X, \vec{\theta}|M_i) d\vec{\theta} \\ &= \int p(X|\vec{\theta}, M_i) \pi(\vec{\theta}|M_i) d\vec{\theta}, \end{aligned} \quad (3.8)$$

where  $\vec{\theta}$  is the set of parameters of model  $M_i$ . Equation (3.8) is thus obtained by averaging the likelihood over the model parameters weighted by their prior probabilities. The probability  $p(X|M_i)$  is thus the probability of the data given the model without assumptions of model parameters. This is called the *marginal likelihood* of the model, also known as the *model evidence* (Bailer-Jones (2011)). The Bayes factor is thus the ratio of marginal likelihoods.

We now express the model posteriors in terms of posterior odds in favor of each model over the constant model  $M_0$ . By doing so we can ignore the normalization constant  $p(X)$ , which can be computationally heavy. We can write the posterior for model  $M_i$  as follows:

$$\begin{aligned} p(M_i|X) &= \frac{p(M_i|X)}{\sum_{j=0}^{\nu} p(M_j|X)} \\ &= \frac{\frac{p(M_i|X)}{p(M_0|X)}}{\sum_{j=0}^{\nu} \frac{p(M_j|X)}{p(M_0|X)}}, \end{aligned} \quad (3.9)$$

where the first equality follows from the normalization rule, i.e.,  $\sum_{j=0}^{\nu} p(M_j|X) = 1$ . Using equation (3.3) we can write equation (3.9) as

$$p(M_i|X) = \frac{\mathcal{O}_{M_i M_0}}{\sum_{j=0}^{\nu} \mathcal{O}_{M_j M_0}}. \quad (3.10)$$

From equation (3.6) and (3.10) we have the following posterior density for the null hypothesis:

$$\begin{aligned} p(H_0|X) &= \frac{\mathcal{O}_{M_0 M_0}}{\sum_{j=0}^{\nu} \mathcal{O}_{M_j M_0}} \\ &= \frac{1}{1 + \sum_{j=1}^{\nu} \mathcal{O}_{M_j M_0}} \end{aligned} \quad (3.11)$$

and from equation (3.7) and equation (3.10) we obtain the posterior density for the alternative hypothesis:

$$p(H_1|X) = \frac{\sum_{j=1}^{\nu} \mathcal{O}_{M_j M_0}}{1 + \sum_{j=1}^{\nu} \mathcal{O}_{M_j M_0}}. \quad (3.12)$$

Also note that the following equality holds:

$$p(H_0|X) + p(H_1|X) = 1.$$

Since we now have found expressions for the posteriors of both hypotheses, let us define the posterior odds  $\mathcal{O}_{H_1 H_0}$  in favor of hypothesis  $H_1$  over hypothesis  $H_0$  (hereafter referred to as the *global posterior odds*). Using equations (3.3), (3.4), (3.11) and (3.12) we can write the global posterior odds as

$$\begin{aligned} \mathcal{O}_{H_1 H_0} &= \sum_{j=1}^{\nu} \mathcal{O}_{M_j M_0} \\ &= \sum_{j=1}^{\nu} \frac{\pi(M_j)}{\pi(M_0)} \mathcal{B}_{M_j M_0}. \end{aligned} \quad (3.13)$$

### Priors distributions

In the GL algorithm, the probability of periodicity is assumed to be equal to the probability of non periodicity. This assumption also holds for the planetary detection algorithm. Therefore we assign equal probabilities of  $\frac{1}{2}$  to the class of non periodic models and to the class of periodic models. Since the class of non periodic models only consists of the constant model  $M_0$ , we have

$$\pi(M_0) = \frac{1}{2}. \quad (3.14)$$

We already introduced  $\nu$  as the total number of periodic models considered. Assuming each periodic model to be equally probable, we have

$$\pi(M_n) = \frac{1}{2\nu}. \quad (3.15)$$

Substituting these priors in equation (3.13) gives:

$$\begin{aligned} \mathcal{O}_{H_1 H_0} &= \sum_{j=1}^{\nu} \frac{\frac{1}{2\nu}}{\frac{1}{2}} \mathcal{B}_{M_j M_0} \\ &= \frac{1}{\nu} \sum_{j=1}^{\nu} \mathcal{B}_{M_j M_0}. \end{aligned} \quad (3.16)$$

We will thus use equation (3.16) to calculate the global posterior odds. Note that by using equation (3.5) we can write equation (3.16) as:

$$\begin{aligned} \mathcal{O}_{H_1 H_0} &= \frac{1}{\nu} \sum_{j=1}^{\nu} \frac{p(X|M_j)}{p(X|M_0)} \\ &= \frac{1}{\nu p(X|M_0)} \sum_{j=1}^{\nu} p(X|M_j). \end{aligned} \quad (3.17)$$

Let us find an expression for the summation in equation (3.17). By Bayes' theorem we can write

$$\begin{aligned} p(X|H_1) &= \frac{p(X)p(H_1|X)}{p(H_1)} \\ &= \frac{p(X)}{p(H_1)} \sum_{j=1}^{\nu} p(M_j|X), \end{aligned} \quad (3.18)$$

where equation (3.18) follows from equation (3.7). Applying Bayes' theorem on  $p(M_j|X)$  in equation (3.18) yields:

$$\begin{aligned} p(X|H_1) &= \frac{p(X)}{p(H_1)} \sum_{j=1}^{\nu} \frac{p(X|M_j)p(M_j)}{p(X)} \\ &= \sum_{j=1}^{\nu} \frac{p(X|M_j)p(M_j)}{p(H_1)}. \end{aligned} \quad (3.19)$$

Since both hypotheses have equal prior probabilities of  $\frac{1}{2}$  and using equation (3.15) we obtain

$$\begin{aligned} p(X|H_1) &= \sum_{j=1}^{\nu} \frac{p(X|M_j) \frac{1}{2^{\nu}}}{\frac{1}{2}} \\ &= \frac{1}{\nu} \sum_{j=1}^{\nu} p(X|M_j). \end{aligned} \quad (3.20)$$

From Bayes' theorem and equation (3.6) we obtain

$$\begin{aligned} p(X|H_0) &= \frac{p(H_0|X)p(X)}{p(H_0)} \\ &= \frac{p(M_0|X)p(X)}{p(H_0)}. \end{aligned} \quad (3.21)$$

Using Bayes' theorem, equation (3.15) and knowing that  $p(H_0) = \frac{1}{2}$  we can write equation (3.21) as:

$$p(X|H_0) = p(X|M_0). \quad (3.22)$$

Substituting equations (3.20) and (3.22) in equation (3.17) and using equation (3.5) yields:

$$\mathcal{O}_{H_1 H_0} = \mathcal{B}_{H_1 H_0}. \quad (3.23)$$

The global posterior odds is thus equal to the Bayes factor  $\mathcal{B}_{H_1 H_0}$  (hereafter referred to as the *global Bayes factor*), which is what we expect since the priors  $\pi(H_0)$  and  $\pi(H_1)$  both equal  $\frac{1}{2}$  resulting in a prior odds equal to 1.

### 3.6.1 Evidence against the null hypothesis

The Bayes factor is a summary of the evidence provided by the data in favor of the alternative hypothesis (transit signal) over the null hypothesis (no transit) (Kass and Raftery (1995)). Sir Harold Jeffreys gave a scale for the interpretation of the strength of the global Bayes factor (Jeffreys (1961), appendix B), see Table 1.

$\mathcal{B}_{H_1 H_0}$	Evidence against $H_0$
1 to 3.2	Not worth more than a bare mention
3.2 to 10	Substantial
10 to 100	Strong
> 100	Decisive

Table 1: Scale for the global Bayes factor.

The scale given by Sir Harold Jeffreys is not an exclusive scale and is no more than a guide to the interpretation of the evidence against the null hypothesis. The scale for the evidence interpretation needs to be settled by consensus among scientists from all disciplines (Evetts et al. (2000)). See for example Table 2 for the scale followed by the Forensic Science Service (FSS). Note that the global Bayes factor is a ratio of marginal likelihoods (see equation (3.5)) in favor of the alternative hypothesis  $H_1$  over the null hypothesis  $H_0$ .

Likelihood ratio	Evidence against $H_0$
> 1 to 10	Limited evidence to support $H_1$
10 to 100	Moderate evidence to support $H_1$
100 to 1000	Moderately strong evidence to support $H_1$
1000 to 10000	Strong evidence to support $H_1$
> 10000	Very strong evidence to support $H_1$

Table 2: Scale for the likelihood ratio followed by the FSS (Evetts et al. (2000)).

Noteworthy is the recommendation given in the AF algorithm about the evidence provided by the global Bayes factor (which equals the global posterior odds, see equation (3.23)). Note that in Paper AF (and in Paper GL) the global posterior odds is called the global odds ratio denoted by  $O_{\text{per}}$ .

*“In Gregory and Loredi (1992), a global odds ratio is calculated ... in order to determine whether there is evidence for a periodic signal. If the global odds ratio is larger than 1, the answer is yes.”*

— Aigrain and Favata (2002) (Paper AF).

From Table 1 and Table 2 we see that when the global Bayes factor is larger than 1, this does not automatically imply good evidence for a periodic signal. For instance, when the global Bayes factor lies within the interval  $[1, 3.2]$ , the evidence is not worth more than a bare mention according to Jeffreys (1961). From Table 2 we see that likelihood ratios smaller than 10 correspond with limited evidence to support  $H_1$ . So in my opinion it is more appropriate to talk about evidence for a periodic signal when larger values for the global Bayes factor are obtained. Also in the GL



algorithm the evidence for the periodic signal is not explained very specifically, although they are more careful by saying that the magnitude of  $O_{\text{per}}$  indicates the strength of the evidence.

*“When  $O_{\text{per}}$  is greater than unity, there is evidence for a periodic signal, the magnitude of  $O_{\text{per}}$  indicating the strength of this evidence.”*

— Gregory and Loredo (1992) (Paper GL).

In order to calculate the global Bayes factor to give statements about which hypothesis is true, we need to calculate the Bayes factors (see equation (3.16) and (3.23))

$$\mathcal{B}_{M_j M_0} = \frac{p(X|M_j)}{p(X|M_0)},$$

which are the ratios of the marginal likelihoods defined in equation (3.8). In the next section we will focus on these likelihood functions.

### 3.7 Likelihood functions

The GL algorithm was initially developed for the detection of X-ray pulsars, which are dominated by Poisson noise. In a subsequent paper, Gregory (1999), the GL algorithm is generalized to the case where time series have a noise sampling distribution of independent Gaussian noise. Although the AF algorithm works with time series with a Poisson noise distribution, the large number of photons in each time bin causes the Poisson noise to be indistinguishable from Gaussian noise (Central Limit Theorem). So the Bayesian planetary detection algorithm (AF algorithm) follows the method in Gregory (1999) to calculate the likelihoods of the models, which are necessary to calculate the global Bayes factors.

In Section 3.2 we defined  $r_j$  as the level in bin  $j$ , where  $j = 1, \dots, m$ . Let us define data points  $x_t$  as our data points of our data  $X$ . In the Bayesian detection algorithm the data set is also folded but in a different way than we did in the box-fitting algorithm. In the GL algorithm, each data point is assigned to a bin number  $j$  by the following:

$$j(t) = \text{int}\{1 + m[(\omega t + \phi) \bmod 2\pi]/2\pi\}. \quad (3.24)$$

Here  $\omega$  is the angular frequency parameter, however we can still use the period parameter  $P$  by the following relation  $\omega = 2\pi/P$ . The AF algorithm for planetary signals makes use of a slightly different type of periodic models, as described in Section 3.3. Therefore, equation (3.24) is replaced in Paper AF by the following:

$$j(t) = \begin{cases} t_{\text{mod}} & \text{if } 0 < t_{\text{mod}} \leq n, \\ 0 & \text{otherwise,} \end{cases} \quad (3.25)$$

where

$$t_{\text{mod}} = \text{int} \left( \frac{(t + P - e) \bmod (P)}{d/n} + 1 \right).$$

Here  $P$ ,  $d$  and  $e$  are the parameters for the periodic model and  $n$  is the number of bins per transit, which refers to the other  $n$  bin parameters  $r_j$ ,  $j = 1 \dots n$ , of the model. In this modified model, every periodic model contains an out-of-transit bin. This bin refers to parameter  $r_0$  which brings

our total of parameters to  $n + 4$ . Note that the constant model,  $M_0$ , contains one constant bin value. This constant bin level is denoted by  $A$ .

Similar to equation (2.2) we can represent any data point  $x_t$  where  $t = 1, \dots, N$  by the following equation:

$$x_t = x_t^p + \epsilon_t, \quad (3.26)$$

where  $x_t^p$  is the value of  $x_t$  predicted by the model and  $\epsilon_t$  is a noise component, which is a zero-mean Gaussian noise term with variance  $\sigma_t^2$  as discussed in Section 1.2. From equation (3.26) we have the following

$$x_t \sim \mathcal{N}(x_t^p, \sigma_t^2). \quad (3.27)$$

Note that in Paper AF heteroscedastic noise is assumed, which is a good approximation for space data, see Section 2.4. Assuming each data point  $x_t$  to be independent, the likelihood can be written as the product of independent Gaussian probability density functions:

$$p(X|P, d, e) = \prod_{t=1}^N \frac{1}{\sigma\sqrt{2\pi}} \exp \left[ -\frac{(x_t - x_t^p)^2}{2\sigma^2} \right]. \quad (3.28)$$

The likelihood given in equation (3.28) needs to be modified to suit the periodic model  $M_n$  with  $n + 1$  bins. The predicted value of  $x_t$  in bin  $j$  is the bin level  $r_j$ . Following the notation in Gregory (1999), we denote the set of  $n + 1$  bin levels  $r_j$  by the vector  $\vec{r}$ . The likelihood for the periodic model  $M_n$  becomes

$$\begin{aligned} p(X|P, d, e, \vec{r}, M_n) &= \left( \frac{1}{\sigma\sqrt{2\pi}} \right)^N \prod_{j=0}^n \prod_{t=1}^{n_j} \exp \left[ -\frac{(x_t - r_j)^2}{2\sigma^2} \right] \\ &= \left( \frac{1}{\sigma\sqrt{2\pi}} \right)^N \prod_{j=0}^n \exp \left[ -\frac{\sum_{t=1}^{n_j} (x_t - r_j)^2}{2\sigma^2} \right], \end{aligned} \quad (3.29)$$

where  $n_j$  is the number of data points in bin  $j$ , following the notation in Paper AF.

In order to calculate the global Bayes factors in equation (3.16), we need to integrate out the parameters in the likelihood in equation (3.29) to obtain the marginal likelihood. From equation (3.8) this marginal likelihood is of the following form:

$$\begin{aligned} p(X|M_n) &= \int dP \int dd \int de \int d\vec{r} \pi(P|M_n) \pi(d|M_n) \\ &\quad \pi(e|M_n) \pi(\vec{r}|M_n) p(D|P, d, e, \vec{r}, M_n). \end{aligned} \quad (3.30)$$

We thus need to assign prior distributions for each of the parameters in order to obtain the marginalized likelihood in equation (3.30). In Section 3.8 we will briefly define the prior distributions used in the AF method.

### 3.7.1 Likelihoods for the constant model

Since the constant model consists of a single bin with a constant level  $A$ , we can omit the parameters  $P$ ,  $d$ , and  $e$  because of the absence of a transit in the constant model. Also the bin parameters  $r_j$  are replaced by the parameter  $A$ . Consequently, the constant model  $M_0$  is a one-parameter

model with parameter  $A$ . Therefore, the likelihood for the constant model is much less complicated than the likelihood for the periodic models. Since all the data points are in the same bin, we can obtain the likelihood for  $M_0$  by setting  $x_t^p = A$  in equation (3.28). By doing so, we obtain the following likelihood for the constant model:

$$p(X|A, M_0) = \left( \frac{1}{\sigma\sqrt{2\pi}} \right)^N \prod_{t=1}^N \exp \left[ -\frac{(x_t - A)^2}{2\sigma^2} \right]. \quad (3.31)$$

The prior distribution assigned to  $A$  in the GL algorithm can also be used in the modified algorithm for planetary signals, since both methods use the same constant model. The prior assigned in the GL algorithm is a flat prior density that is constant from  $A = 0$  to some upper limit  $A_{\max}$ . The normalized prior density for  $A$  is then

$$\pi(A|M_0) = \frac{1}{A_{\max}}. \quad (3.32)$$

From equations (3.8), (3.31) and (3.32), the marginal likelihood for the constant model is

$$\begin{aligned} p(X|M_0) &= \int \pi(A|M_0) p(X|A, M_0) dA \\ &= \int \frac{1}{A_{\max}} \left( \frac{1}{\sigma\sqrt{2\pi}} \right)^N \prod_{i=1}^N \exp \left[ -\frac{(x_t - A)^2}{2\sigma^2} \right] dA. \end{aligned} \quad (3.33)$$

### 3.8 Prior probabilities for periodic model parameters

In the introduction of this chapter we mentioned the different schools of thoughts in the Bayesian inference: “subjectivism” and “objectivism”. Subjectivists assign prior distributions to parameters which reflect their own subjective opinion about the parameter. In some cases this may be possible. However, in complicated problems where the number of parameters is high and the prior information about the problem is relatively scarce, subjective priors seem to be impractical (Wasserman (2004), p.213). The planetary transit detection problem is a complicated problem and the number of parameters for some periodic models  $M_n$  can be high when  $n$  increases.

The priors chosen in the AF algorithm are thus *non-informative* priors. Therefore we will not pay much attention to the prior allocation. An obvious candidate for a non-informative prior is a *flat prior* (Wasserman (2004), p.213)

$$\pi(\theta) \propto \text{constant},$$

like the prior for  $A$  in equation (3.32). The priors assigned to parameters  $d$ ,  $e$  and bin levels  $r_j$  are also non-informative flat priors similar to the prior assigned to parameter  $A$  for the constant model. For example, the bin level parameters  $r_j$  have a flat prior density in the range of  $r_{\min}$  to  $r_{\max}$ :

$$\pi(r_j|M_n) = \frac{1}{r_{\max} - r_{\min}}. \quad (3.34)$$

As we have seen in Section 3.7 we can work with an angular frequency parameter  $\omega$  rather than the period parameter  $P$ . We therefore want a form-invariant prior with respect to parameterization. That is, an investigator working in terms of period  $P$  will reach the same conclusions as

an investigator working in terms of  $\omega$  (Gregory (1999)). A flat prior will not suffice, since these priors are not *transformation invariant* (Wasserman (2004), p.213).

Sir Harold Jeffreys came up with a rule for creating priors which are transformation invariant. The *Jeffreys' prior* is defined as follows: (Wasserman (2004), p.213)

$$\pi(\theta) \propto \sqrt{I(\theta)},$$

where  $I(\theta)$  is the *Fisher information function*. The AF algorithm used the same Jeffreys' prior as the prior used in the GL algorithm:

$$p(\omega|M_n) = \frac{1}{\omega \ln(\omega_{\text{hi}}/\omega_{\text{lo}})},$$

where  $(\omega_{\text{lo}}, \omega_{\text{hi}})$  is a prior range for  $\omega$ . The fact that the AF algorithm works with parameter  $P$  instead of  $\omega$  is thus not a problem. For more information about Jeffreys's priors see Jeffreys (1946).

### 3.9 Calculations for the global Bayes factor

From Section 3.6.1 we need to bear in mind that we need to calculate the Bayes factors

$$\mathcal{B}_{M_j M_0} = \frac{p(X|M_j)}{p(X|M_0)}, \quad (3.35)$$

where  $j = 1, \dots, \nu$ , to calculate the global Bayes factor  $\mathcal{B}_{H_1 H_0}$  in order to make statements about under which hypothesis our data has been arisen. From Table 1 we see that the evidence against  $H_0$  becomes stronger when  $\mathcal{B}_{H_1 H_0}$  becomes larger. Of course in the ideal situation we obtain  $\mathcal{B}_{H_1 H_0} > 100$  which indicates decisive evidence against  $H_0$  so that we can conclude that our data  $X$  contains planetary transits. In Section 3.7 (equation (3.30) and (3.33)) we found expressions for the marginal likelihoods in equation (3.35), so we are able to compute the global Bayes factors. However, in this section we will argue that these calculations are not that simple to compute and that these heavy calculations affect the speed of the algorithm.

#### 3.9.1 Disadvantage of Bayesian algorithm with respect to detection purposes

In Section 1.1.1 we discussed that because of the low transit probability, the telescopes need to observe a very large number of stars. Consequently the detection algorithms need to be very efficient and fast, i.e., the algorithms need a low computation time. A big obstacle in the investigated Bayesian algorithm is the calculation of the marginal likelihood for each model.

We see in equation (3.30) that we need to calculate multiple integrals in order to calculate the marginal likelihood. Each integral represents a parameter, which needs to be integrated out, and this number of integrals increases when the number of bin parameters in the model increases. These integrals cannot be computed analytically, and thus need to be computed numerically, which are quite intensive computations. Since we are focused on *detection purposes* only, it is obvious that this kind of computations for each data set of one of the observed stars (and for multiple models per data set) is not optimal.

So compared to the box-fitting algorithm the Bayesian algorithm is less suitable to operate as a detection algorithm since the box-fitting algorithm does not contain heavy computations like the Bayesian algorithm. The box-fitting algorithm also makes use of just one model in contrast to the Bayesian algorithm which uses different periodic models. The Bayesian algorithm thus

needs to be simplified in order to become an efficient transit detection algorithm. In a subsequent paper (Aigrain and Irwin (2004)) the Bayesian algorithm has been further developed into a more efficient algorithm for detection purposes. Multiple simplifications are made to reduce the computational requirements. We will further investigate this simplified algorithm in Chapter 4.

### 3.9.2 Parameter estimation

For detection purposes the Bayesian algorithm is thus not the most optimal algorithm. However, the introduction of this algorithm has not been in vain. The detection process is the first (and of course a very important) step in exoplanetary research. When an exoplanet is *detected*, one can subsequently focus on the estimation of the planet's parameters. By doing so, one can find answers to questions as "is the exoplanet a terrestrial Earth-like planet?" and "could the exoplanet support water in its liquid state?"<sup>1</sup> These are the questions that matter and were the initial motivation for the launch of the Kepler telescope in 2009.

The shape of the transit carries a lot of information about the planetary parameters (Sackett (1999)). So in order to make good estimations of these parameters, we need to approximate this transit shape very accurately. From Figure 3.2 we see that the Bayesian algorithm is able to approximate the shape of the light curve by using models with a large number of bins per period. So a high number of bin parameters yields a accurate approximation of the transit.

Next to accurate estimates of the bin parameters, good estimates of the other parameters such as the period  $P$  are also required in order to estimate the transit shape. We now argue how the Bayesian algorithm provides parameter estimates using the posterior distributions of the parameters. When the global Bayes factor is large enough to conclude the presence of an orbiting exoplanet, the second part of the algorithm is the estimation of the parameters. We can now use the marginal likelihood to derive posterior distributions for the model parameters using Bayes' theorem again. This is done first for the  $n$  bin periodic model and later marginalization over  $n$  follows (Gregory (1999)). For instance, the posterior density for the period  $P$  is:

$$p(P|X, M_n) = \frac{\pi(P|M_n)p(X|P, M_n)}{p(X|M_n)}, \quad (3.36)$$

where we see that the marginal likelihood plays the role of a normalization constant. We discussed the allocation of the priors in Section 3.8 and the likelihood in the numerator of equation (3.36) is obtained by integrating out all nuisance parameters of the product of the likelihood in equation (3.29) with the priors of the nuisance parameters.

Strictly speaking, we do not obtain an estimation for the parameter  $P$ . The Bayesian approach to the parameter estimation problem is the posterior distribution in equation (3.36). This results in a rigorous basis for quantifying uncertainties for model parameters, which is an advantage of Bayesian inference (Ford (2006)) with respect to frequentist inference where estimates are single points in the parameter space. It is often useful to summarize the posterior in terms of a "best-fit" value. Potential summarizing best-fit values are the *posterior mode* (most probable value of  $P$ ) or

---

<sup>1</sup>In order to support water in its liquid state, the exoplanet must fall in the star's *habitable zone*. This zone is also referred to as the "Goldilocks zone". This name comes from the fairy tale *Goldilocks and the three bears* wherein Goldilocks tastes three bowls of porridge. The first bowl is too cold, the second bowl is too hot and the last bowl has the perfect temperature. The same logic holds for the exoplanet: the distance towards its host star must be perfect so that the surface temperature supports liquid water and thus life (as we know it).

the *posterior mean* (Paper GL)

$$\langle P \rangle = \int p(P|X, M_n) P dP.$$

*Credible regions* (analogues of the frequentist confidence interval) can also be obtained from the posterior density (Paper GL).

When certain exoplanets are detected, one can thus use the second part of the Bayesian algorithm to obtain posterior distributions for the model parameters. The interested reader is referred to Paper GL and Gregory (1999) for more information about the Bayesian parameter estimation.

### 3.9.3 MCMC and radical velocity

In equation (3.36) we see that next to the marginal likelihood in the denominator, we also need to calculate  $p(X|P, M_n)$ , which is an almost fully marginalized likelihood. In Section 3.9.1 we discussed that these marginal likelihoods are extremely difficult to compute, particularly when the number of parameters is large. For the detection purposes these calculations are too expensive, but after the detection process we have the privilege to focus on a single light curve. So instead of thousands of calculations, a few marginal likelihood computations are executable. However, the numerical calculation of the integrals are of course still very difficult. For a long time, this computational burden has limited Bayesian methods to a small number of relatively simple problems. The brute force evaluation of the integrals is even today impractical for many complex problems (Ford (2006)). Therefore, efficient algorithms are developed to evaluate the integrals. The method of *Markov chain Monte Carlo* (MCMC) has become an increasingly common tool for evaluating the Bayesian integrals. For more information about MCMC see Ford (2005) and references therein.

The exoplanets detected during the detection process are called *planet candidates* which need to be confirmed by follow-up measurements. Planet candidates detected by the transit method are often confirmed by radial velocity observations. Bayesian inference plays a major role in these follow-up measurements, which agrees with the following quote:

*“Astronomers searching for the small signals induced by planets inevitably face significant statistical challenges. Bayesian inference has the potential of improving the interpretation of existing observations, the planning of future observations and ultimately inferences concerning the overall population of planets.”*

— P.C. Gregory (2007).

Bayesian inference is potentially more suitable for parameter estimation and follow-up measurements for planet candidates. We see in the literature that Bayesian inference is mainly used in the radial velocity method. See e.g. the works of Gregory (2005, 2007) for a Bayesian re-analysis of earlier detected exoplanets using the radial velocity method. More Bayesian inference for radial velocity observations: <http://www.phas.ubc.ca/~gregory/gregory.html>

## 4 Simplified “Bayesian” detection algorithm

### 4.1 Introduction

In this chapter we consider the transit search algorithm which was presented in a paper following Paper GL (Aigrain and Irwin (2004), hereafter referred to as Paper AI) as a simplified version of the Bayesian algorithm described in Chapter 3. This simplification has been made in order to make the algorithm more efficient for detection purposes. We will see in this chapter that this new algorithm (hereafter referred to as AI algorithm) needs less computational requirements and thus is a more suitable algorithm for detection purposes only. In the first section we explain the simplifications made to the Bayesian algorithm. We will see that the new algorithm is based on a maximum-likelihood approach and loses the Bayesian interpretation because of the simplifications made to the Bayesian algorithm. In the last section we will compare the new algorithm to an extant algorithm published before Paper AI, the box-fitting algorithm. We end this chapter by showing that these two algorithms are equivalent.

### 4.2 Defining the algorithm

The obstacle in the Bayesian algorithm for the detection process is the calculation of the marginal likelihoods, which can exist of a large number of integrals depending on the number of parameters. Especially the number of bin parameters are large for some models. The first simplification made to the Bayesian algorithm is excluding the use of the bin levels  $r_j$  as free parameters. In Paper AI it is demonstrated that the optimal bin levels for each model can be directly determined from the data using likelihood maximization given the other model parameters  $P$ ,  $d$  and  $e$ . Also the AI algorithm does not make use of Bayesian prior distributions, since the lack of information available on exoplanetary characteristics. Another counterargument given in Paper AI is that the Bayesian priors used in Chapter 3 were identical to the priors used for the detection of X-ray pulsars (Paper GL), rather than specifically optimized for exoplanetary transit searches. In my opinion the latter counterargument is inappropriate since the priors used for X-ray pulsars were non-informative priors. Also the lack of exoplanetary information should make no difference when using non-informative priors. Because the Bayesian priors are dropped and fixed trial parameters are used, the AI algorithm loses the Bayesian interpretation. Consequently, the AI algorithm avoids the problem of calculating computationally intensive marginal likelihoods.

The model  $M_n$  has  $n + 4$  parameters in the Bayesian algorithm:  $n + 1$  bin levels, period  $P$ , duration  $d$  and epoch  $e$ . In the AI algorithm the set of free model parameters consists of four parameters:  $P$ ,  $d$ ,  $e$  and the number of bins per period  $n$ . We will see that the bin levels can be fully defined by the data. The AI algorithm treats the parameters as fixed constants instead of random variables, hence the statistical inference in the simplified method is a frequentist approach. For a given number of bins  $n$ , the AI algorithm finds the best set of parameters  $P$ ,  $d$ ,  $e$  and bin levels  $r_j$  by searching for the model most likely to have given rise to the observed data, i.e., the model which yields the maximum likelihood.

#### 4.2.1 Maximum-likelihood approach

Since the zero-mean Gaussian noise assumption made in Section 1.2 still holds for the AI algorithm, we recall the distribution given in (3.27):

$$x_t \sim \mathcal{N}(x_t^p, \sigma_t^2),$$

where  $x_t^p$  is the value of  $x_t$  predicted by the model. Assuming independency of the data points we have the following likelihood function:

$$L(x_t^p, \sigma_t) = \prod_{t=1}^N \frac{1}{\sigma_t \sqrt{2\pi}} \exp \left[ -\frac{(x_t - x_t^p)^2}{2\sigma_t^2} \right]. \quad (4.1)$$

Let us rewrite this likelihood as follows:

$$\begin{aligned} L(x_t^p, \sigma_t) &= \left( \frac{1}{\sqrt{2\pi}} \right)^N \prod_{t=1}^N \frac{1}{\sigma_t} \prod_{t=1}^N \exp \left[ -\frac{(x_t - x_t^p)^2}{2\sigma_t^2} \right] \\ &= \left( \frac{1}{\sqrt{2\pi}} \right)^N \prod_{t=1}^N \frac{1}{\sigma_t} \exp \left[ \sum_{t=1}^N -\frac{(x_t - x_t^p)^2}{2\sigma_t^2} \right] \\ &= \left( \frac{1}{\sqrt{2\pi}} \right)^N \prod_{t=1}^N \frac{1}{\sigma_t} \exp \left( -\frac{\chi^2}{2} \right), \end{aligned} \quad (4.2)$$

where

$$\chi^2 = \sum_{t=1}^N \frac{(x_t - x_t^p)^2}{\sigma_t^2}. \quad (4.3)$$

So likelihood maximization is equivalent to  $\chi^2$  minimization, since the noise properties  $\sigma_i$  are assumed to be known (Paper AI). The AI algorithm maximizes the likelihood for each model  $M_n$ . Note that the influence of the parameters  $P, d$  and  $e$  are expressed in the expression of  $j(t)$  which represents the bin allocation in equation (3.25). In Paper AI  $\chi_j^2$  is defined as the contribution of each bin  $j$  to the overall  $\chi^2$  where  $j = 1, \dots, n$ :

$$\chi_j^2 = \sum_{t=1}^{n_j} \frac{(x_t - r_j)^2}{\sigma_t^2}, \quad (4.4)$$

where  $n_j$  is the number of data points in bin  $j$ . Equation (4.3) can subsequently be rewritten as

$$\chi^2 = \sum_{j=0}^n \chi_j^2,$$

where bin  $j = 0$  corresponds with the out-of-transit bin. The bin parameters  $r_j$ , which replace the variables  $x_i^p$  in equation (4.3), can be fully obtained by the data by maximizing the likelihood, i.e., minimizing  $\chi^2$ . So for trial parameters  $P, d, e$  and  $n$  the AI algorithm aims to find the bin levels  $r_j$  which minimize the overall  $\chi^2$ . Let us summarize this process in the model  $\mathcal{M}$ :

$$\mathcal{M} = \{ \chi^2, (P, d, e, n, \vec{r}) \in \Theta \},$$



where  $\Theta$  is the parameter space for parameter combinations.

So for given trial parameters  $P, d, e$  and  $n$  the algorithm needs to minimize  $\chi^2$ . The AI algorithm defines  $\tilde{r}_j$  as the value of the bin level  $r_j$  that minimizes  $\chi_j^2$ . This value can be obtained by setting the derivative  $\partial\chi_j^2/\partial r_j$  equal to zero which yields:

$$2 \sum_{t=1}^{n_j} \frac{r_j - x_t}{\sigma_t^2} = 0,$$

and hence

$$\tilde{r}_j = \frac{\sum_{t=1}^{n_j} x_t \sigma_t^{-2}}{\sum_{t=1}^{n_j} \sigma_t^{-1}}. \quad (4.5)$$

By substituting equation (4.5) into equation (4.4) we can write the minimized value of  $\chi_j^2$  as

$$\widetilde{\chi_j^2} = \sum_{t=1}^{n_j} \frac{(x_t - \tilde{r}_j)^2}{\sigma_t^2}. \quad (4.6)$$

Expanding equation (4.6) gives:

$$\widetilde{\chi_j^2} = \sum_{t=1}^{n_j} \frac{x_t^2}{\sigma_t^2} - 2\tilde{r}_j \sum_{t=1}^{n_j} \frac{x_t}{\sigma_t^2} + \tilde{r}_j^2 \sum_{t=1}^{n_j} \frac{1}{\sigma_t^2}.$$

Rewriting equation (4.5) yields

$$\sum_{t=1}^{n_j} \frac{x_t}{\sigma_t^2} = \tilde{r}_j \sum_{t=1}^{n_j} \frac{1}{\sigma_t^2}, \quad (4.7)$$

so that

$$\widetilde{\chi_j^2} = \sum_{t=1}^{n_j} \frac{x_t^2}{\sigma_t^2} - \tilde{r}_j^2 \sum_{t=1}^{n_j} \frac{1}{\sigma_t^2}.$$

We can now define the overall minimized  $\chi^2$  by

$$\begin{aligned} \widetilde{\chi^2} &= \sum_{j=0}^n \widetilde{\chi_j^2} \\ &= \sum_{j=0}^n \left[ \sum_{t=1}^{n_j} \frac{x_t^2}{\sigma_t^2} - \tilde{r}_j^2 \sum_{t=1}^{n_j} \frac{1}{\sigma_t^2} \right] \\ &= \sum_{t=1}^N \frac{x_t^2}{\sigma_t^2} - \sum_{j=0}^n \left[ \tilde{r}_j^2 \sum_{t=1}^{n_j} \frac{1}{\sigma_t^2} \right]. \end{aligned} \quad (4.8)$$

The optimal parameter set for model  $\mathcal{M}$  thus corresponds with the parameter set which yields the absolute minimum of the calculated  $\widetilde{\chi^2}$ . Note that the first term in equation (4.8) is independent of the parameter set, therefore the algorithm only needs to calculate the second term. The new algorithm thus builds further on the models introduced in the Bayesian algorithm, however the computational requirements have been significantly reduced. By directly determining the bin levels using likelihood maximization no computationally heavy integrals need to be calculated, which benefit the detection process.

#### 4.2.2 Final reducing modification

A final simplification made in Paper AI is to reduce the number of bins per transit to one. The Bayesian algorithm can approximate the transit dip with multiple bins. The approximation of the shape of the transit becomes more accurate when the number of bins per transit increases. As we discussed in Chapter 2, the shape of the transit is not of interest for detection purposes. Therefore a single in-transit bin provides enough information, see Figure 4.1. The reduction to one in-transit bin reduces computational requirements and the loss in detection efficiency is negligible. In Paper AI the out-of-transit and in-transit bin are defined as  $O$  and  $I$  respectively.

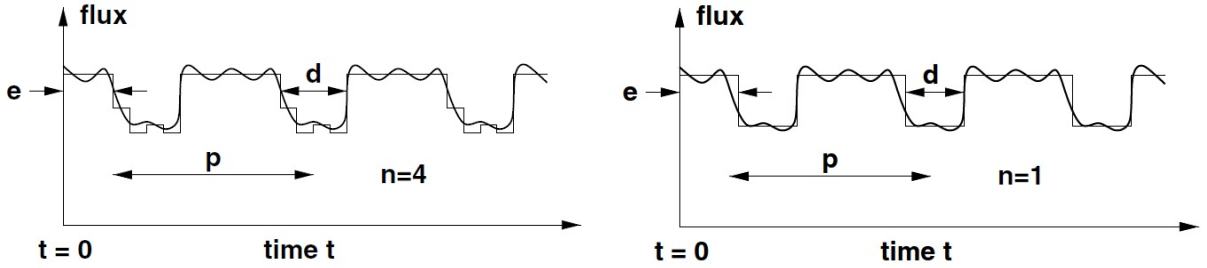


Figure 4.1: Illustration of the type of models used in the Bayesian algorithm (left) which can approximate the transit with multiple bins and the model used in the AI algorithm (right). Figure by Aigrain and Irwin (2004).

Equation (4.8) can thus be rewritten as

$$\widetilde{\chi^2} = \sum_{t=1}^N \frac{x_t^2}{\sigma_t^2} - \widetilde{r}_O^2 \sum_{t \in O} \frac{1}{\sigma_t^2} - \widetilde{r}_I^2 \sum_{t \in I} \frac{1}{\sigma_t^2}. \quad (4.9)$$

Given that the transits are short (typically less than 5%, Section 1.2), in Paper AI  $\widetilde{r}_O^2 \approx \widetilde{x}^2$  is assumed. Here  $\widetilde{x}^2$  is the weighted mean of the entire light curve. Substituting this approximation in equation (4.9) gives:

$$\widetilde{\chi^2} \approx \sum_{t=1}^N \left[ \frac{x_t^2}{\sigma_t^2} - \frac{\widetilde{x}^2}{\sigma_t^2} \right] - \widetilde{r}_I^2 \sum_{t \in I} \frac{1}{\sigma_t^2}. \quad (4.10)$$

Note that the first two terms in equation (4.10) are constant. Minimization of  $\widetilde{\chi^2}$  can thus be obtained by maximizing the last term in equation (4.10), which can be expanded by equation (4.7) as

$$\widetilde{r}_I^2 \sum_{t \in I} \frac{1}{\sigma_t^2} = \left( \sum_{t \in I} \frac{x_t}{\sigma_t^2} \right)^2 \left( \sum_{t \in I} \frac{1}{\sigma_t^2} \right)^{-1}. \quad (4.11)$$

### 4.3 Similarity with box-fitting algorithm

The simplifications to the Bayesian algorithm lead to an algorithm based on a frequentist approach. Instead of using multiple models with different parameter values the use of one model

with a single in-transit bin has been used for the sake of detection purposes. The model parameters for the simplified algorithm are thus  $\tilde{r}_O$ ,  $\tilde{r}_I$ ,  $P$ ,  $d$  and  $e$ . This parameter set is the same parameter set used in the model for the box-fitting algorithm, where  $\tilde{r}_O$  and  $\tilde{r}_I$  correspond with  $H$  and  $L$  respectively and parameter  $d$  equals parameter  $q$ . To characterize the quality of the fit, the box-fitting algorithm determines the Signal Residue given in equation (2.11). We see that this Signal Residue is similar to the expression given in equation (4.11) (differences between the two arise by the use of weights in the box-fitting algorithm). The quantity  $s^2/r$  in the Signal Residue coincides with the first term in equation (4.11) and also the Signal Residue is divided by a summation over noise terms in the in-transit interval.

The difference between the box-fitting algorithm and the AI algorithm is the statistical method to determine the levels of  $(H, L)$  and  $(\tilde{r}_O, \tilde{r}_I)$  respectively. The box-fitting algorithm makes use of the method of weighted least squares and the AI algorithm makes use of the maximum-likelihood method. Although these two methods have different names, we conclude this thesis by showing that these methods used in the algorithms are equivalent.

**Fact.** *The maximum-likelihood estimator is equivalent to the weighted least squares estimator under the assumption of normality.*

**Proof.** By equation (3.26) we have

$$x_t = x_t^p + \epsilon_t,$$

where  $\epsilon_t \sim \mathcal{N}(0, \sigma_t^2)$ , i.e.,

$$x_t \sim \mathcal{N}(x_t^p, \sigma_t^2).$$

The likelihood function  $L(x_t^p, \sigma_t)$  is given in equation (4.1). Since the logarithm is a monotonic function, maximizing the log likelihood  $l((x_t^p, \sigma_t))$  yields the same estimator as the maximum likelihood estimator (Rice (2007), p.268). The log likelihood is

$$\begin{aligned} l(x_t^p, \sigma_t) &= \log \left[ \prod_{t=1}^N \frac{1}{\sigma_t \sqrt{2\pi}} \exp \left[ -\frac{(x_t - x_t^p)^2}{2\sigma_t^2} \right] \right] \\ &= \sum_{t=1}^N \log \left[ \frac{1}{\sigma_t \sqrt{2\pi}} \exp \left[ -\frac{(x_t - x_t^p)^2}{2\sigma_t^2} \right] \right] \\ &= \sum_{t=1}^N \log \left( \frac{1}{\sigma_t \sqrt{2\pi}} \right) - \frac{1}{2} \sum_{t=1}^N \frac{1}{\sigma_t^2} (x_t - x_t^p)^2. \end{aligned} \quad (4.12)$$

The maximum likelihood estimator can thus be found by minimizing the log likelihood, i.e., minimizing the second summation in equation (4.12). The weighted least squares estimator can be found by minimizing the weighted residual sum of squares, which is exactly the second summation in equation (4.12).  $\square$

## Conclusion

In this thesis we explored three different exoplanetary transit detection algorithms. The main goal in this thesis was to investigate which algorithm benefits the detection purposes most. Due to geometrical alignment requirements the transit probability is extremely small. In order to enlarge the probability of observing a transit, one has to simply extend the number of observed stars. The detection algorithms thus need to be fast and effective to be able to deal with this large number of observations. We therefore emphasized on computational requirements. Note that the algorithms are only optimized for light data embedded in white Gaussian noise. In this thesis we assume that we obtain data sets filtered to remove red correlated noise since we do not discuss these pre-filtering processes in this thesis.

The first algorithm examined is the box-fitting algorithm in Chapter 2, which is the most common approach in planet searches. This algorithm searches for planetary transits by fitting multiple box-shaped functions to the light curves. The different box-shaped functions are obtained by certain trial parameters for the period, fractional transit length and epoch. Each fit was subsequently evaluated using a least squares estimator. The effectiveness of the box-fitting algorithm relies on the simple shape of the box-shaped functions. Instead of fitting functions which approximates the transit shape, limb darkening effects and gradual ingress and egress phases of the transit are ignored resulting in a box-approximation of the transit. These simplified assumptions are not expected to affect the results of the search. A consequence of these shape simplifications is the loss of important information about the planetary parameters which are carried in the shape of the planetary transit. However, this loss of parameter information is justified as long as the main purposes are the detection purposes. The significant detection threshold found in Paper I is a signal-to-noise ratio larger than 6. In an open problem we suggested to test this threshold using a signal detection analysis.

The Bayesian algorithm in Chapter 3 detects planetary transits using Bayesian inference. In this statistical approach, all unknown quantities contained in a probability model for the data are treated as random variables. So instead of fixed values as parameter estimations as in the box-fitting algorithm, the Bayesian algorithm determines probability distributions for the parameters and models. The algorithm detects periodic transits by Bayesian model comparison to compare a constant model (no transit) to a class of periodic models (with transits). There is evidence for a periodic transit signal if the global Bayes factor is larger than 1 according to Paper AF. In my opinion it is more appropriate to talk about relevant evidence against the constant model when much larger values of the global Bayes factor are obtained. In order to calculate the global Bayes factors, marginal likelihoods need to be calculated which are extremely difficult to compute. This does not benefit the detection purposes and is thus an obstacle for the Bayesian algorithm. We argued that seeing the large number of observed stars, this algorithm is not optimal to function as a detection algorithm. However, the Bayesian algorithm is capable to determine accurate estimations of the planetary parameters which is of great importance after the detection process.

For the sake of detection purposes the Bayesian algorithm has been modified into a simplified detection algorithm. These simplifications lead to an algorithm based on frequentist inference. By directly determining the bin parameters from the data using the maximum-likelihood method, the algorithm obviates the difficult marginal likelihood computations which thus benefits the detection purposes. A second simplification results in the fact that the simplified algorithm uses trial parameters for the period, epoch and duration in order to fit different step functions with

one lower level: a box-shaped function. The high and low level of the box-shaped function were obtained by maximum-likelihood estimators. Since the maximum-likelihood estimator and the weighted least squares estimator are equivalent under the assumption of normality, the simplified algorithm is mathematically identical to the box-fitting algorithm.

The work in this thesis emphasized the position of the box-fitting algorithm as the most common algorithm used in exoplanetary searches. The effectiveness of the algorithm relies on the simple shape of the box-shaped function, which is well capable of detecting planetary transits and thus causes the algorithm to be a very highly efficient algorithm.



## References

- Aigrain, S., Favata, F. (2002), Bayesian detection of planetary transits: A modified version of the Gregory-Loredo method for Bayesian periodic signal detection, *A&A*, 395, 625-636. **[Paper AF]**
- Aigrain, S., Irwin, M. (2004), Practical planet prospecting, *MNRAS*, 350, 331-345. **[Paper AI]**
- Bailer-Jones, C. A. L. (2011), Bayesian time series analysis of terrestrial impact cratering, *MNRAS*, 418, 2111-2112.
- Berger, J. (2006), The case for objective Bayesian analysis, *Bayesian Anal.*, Volume 1, Number 3, 385-402.
- Bijma, F., Jonker, M., Van der Vaart, A. (2013), Inleiding in de statistiek, *Epsilon uitgaven*, Utrecht.
- Deeming T. J. (1975), Fourier analysis with unequally-spaced data, *Ap & SS* 36, 137-158.
- Egan, J.P., (1975), Signal detection theory and ROC analysis, *Series in Cognition and Perception*. Academic Press, New York.
- Evet, I. W., Jackson, G., Lambert, J. A., McCrossan, S. (2000), The impact of the principles of evidence interpretation on the structure and content of statements, *Sci Justice* 40 (4): 233-9.
- Favata and the Eddington Science Team (2000), Eddington assessment study report, *Tech. Rep. ESA - SCI (2000)*, 8, ESA.
- Fawcett, T. (2006), An introduction to ROC analysis, *Journal Pattern Recognition Letters - Special issue: ROC analysis in pattern recognition*, Volume 27, Issue 8, 861-874 .
- Ford, E. B. (2005), Quantifying the uncertainty in the orbits of extrasolar planets, *AJ*, 129, 1706 - 1717.
- Ford, E. B. (2006), Improving the efficiency of Markov Chain Monte Carlo for analyzing the orbits of extrasolar planets, *ApJ*, 642, 505-522.
- Gregory, P. C. (1999), Bayesian periodic detection. I. Analysis of 20 years of radio flux measurements of the X-ray binary LS I 61°303, *ApJ*, 520, 361-375.
- Gregory, P. C. (2005), A Bayesian analysis of extrasolar planet data for HD 73526, *ApJ*, 631, 1198.
- Gregory, P. C. (2007), A Bayesian re-analysis of HD 11964: evidence for three planets, in *Bayesian inference and maximum entropy methods in science and engineering: 27th international workshop*, Knuth, H. L., Caticha, A., Center, J. L., Giffin, A., Rodríguez, C. C., eds, *AIP Conference Proceedings*, 954, 307-314.
- Gregory, P. C., Loredo, J. (1992), A new method for the detection of a periodic signal of unknown shape and period, *ApJ*, 398, 146 - 168. **[Paper GL]**
- Jaynes, E. T. (2003), Probability theory: the logic of science, *Cambridge university press*, New York.
- Jeffreys, H. (1935), Some tests of significance, treated by the theory of probability, *Proceedings of the Cambridge Philosophy Society*, 31, 203 - 222.
- Jeffreys, H. (1946), An invariant form for the prior probability in estimation problems, *Proceedings of the Royal Society of London. Series A., mathematical and physical sciences*, Vol. 186, No. 1007, 453 - 461.
- Jeffreys, H. (1961), Theory of probability, *Clarendon Press*, third edition, Oxford.
- Kass, R. E., Adrian, E. R. (1995), Bayes factors, *Journal of the American Statistical Association*, Vol. 90, No. 430, 773 - 795.
- Kovács, G., Zucker, S., Mazeh, T. (2002) , A box-fitting algorithm in the search for periodic transits, *A&A*, 391, 369 - 377. **[Paper I]**

- McQuillan, M. C., Aigrain, S., Roberts, S. (2012), Statistics of stellar variability from Kepler I. Revisiting Quarter 1 with an astrophysically robust systematics correction, *A&A*, 539, A137.
- Perryman, M. (2011), The exoplanet handbook, *Cambridge university press*, New York.
- Pont, F., Zucker, S., Queloz, D. (2006), The effect of red noise on planetary transit detection, *MNRAS*, 373, 231-242.
- Rice, J. A. (2007), Mathematical statistics and data analysis, *Brooks/Cole*, Berkeley.
- Sackett, P.D. (1999), Planets Outside the Solar System: Theory and Observations (NATO-ASI), eds. Mariotti, J. M., Alloin, D., *Kluwer Academic Publisher*, 189.
- Seager, S., Mallén-Ornelas, G. (2003), A unique solution of planet and star parameters from an extrasolar planet transit light curve, *ApJ*, 85, 1038-1055.
- Smith, J. C., Stumpe, M. C., Van Cleve, J. E., et al. (2012), Kepler presearch data conditioning II - A Bayesian approach to systematic error correction, *arXiv:1203.1383*.
- Snellen, I. A. G., de Mooij, E. J. W., Albrecht, S. (2009), The changing phases of extrasolar planet CoRoT - 1b, *Nature*, 459, 543-545.
- Stellingwerf, R. F. (1978), Period determination using phase dispersion minimization, *ApJ*, 224, 953-960.
- Torres, G., Winn, J. N., Holman, M. J. (2008), Improved parameters for extrasolar transiting planets, *ApJ*, 677, 1324 -1342.
- Wakefield, J. (2013), Bayesian and frequentist regression methods, *Springer*, New York.
- Wasserman, L. (2004), All of statistics: a concise course in statistical inference, *Springer*, New York.
- Winn, J. N. (2009), Measuring accurate transit parameters, *IAU Symposium, volume 253 of IAU Symposium*, 99-109.