Title
Detection Methods for Astronomical Time Series

Author
Coehlo, Nathan Kirk

Publication Date
2010

Peer reviewed|Thesis/dissertation

Detection Methods for Astronomical Time Series

by

Nathan Kirk Coehlo

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Statistics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:
Professor John Rice, Chair
Professor Noureddine El Karoui
Professor Imke de Pater

Spring 2010

Detection Methods for Astronomical Time Series

Abstract

Detection Methods for Astronomical Time Series

by

Nathan Kirk Coehlo
Doctor of Philosophy in Statistics

University of California, Berkeley

Professor John Rice, Chair

In this dissertation I explore and develop statistical methodology motivated by the scientific goals and engineering constraints of the Taiwan-America-Occultation-Survey (TAOS), a large astronomy survey dedicated to understanding populations in our outer solar system. I derive minimal conditions under which detection procedures are valid and robust, discuss the filtering and diagnostics necessary to meet those conditions, and explain how detection results lead to inference about population parameters. I also develop and compare several parametric and non-parametric detection methods for use with TAOS data, and also for the proposed TAOS II survey, which will have higher-resolution data and lower noise levels.

Contents

1 Background and Motivation **1**
1.1 Scientific Background . . . 1
1.1.1 Trans-Neptunian Objects . . . 1
1.1.2 Population Quantities . . . 2
1.1.3 Direct Observations . . . 3
1.1.4 Indirect Observations: Occultations . . . 4
1.2 The TAOS Project . . . 4
1.2.1 Photometry and Zipper Mode . . . 5
1.3 Occultation Patterns and Parameterization . . . 5
1.4 Models for Occultation Data . . . 8
1.5 Statistical Questions . . . 9

2 The Rank Product Method **13**
2.1 Motivation for the Method . . . 13
2.2 The Rank Product Method . . . 13
2.3 Calculation of P-Values . . . 14
2.3.1 A Simple Example . . . 14
2.4 Continuous Approximation: Fisher's Method . . . 14
2.5 Quality of the Continuous Approximation . . . 15
2.6 Exact Calculation of Tail Probabilities . . . 15
2.6.1 Examples . . . 17
2.6.2 Important Cutoffs . . . 17
2.7 Test at time points or over the whole series ? . . . 18
2.8 Weaker Assumptions for the Rank Product Distribution . . . 18
2.8.1 Exact Results . . . 19
2.8.2 Asymptotic Results . . . 20

3 Efficiency **24**
3.1 Expected Number of Detections . . . 24
3.1.1 Simple Illustration of Efficiency . . . 25
3.1.2 Proper Evaluation of f . . . 26
3.1.3 Relative Velocity and Opposition Angle Geometry . . . 27
3.1.4 Proof of lemma 3.1.4 . . . 28
3.2 Inference about population models . . . 30

3.2.1 TAOS Results . . . 30
3.3 Estimation of $\Omega_e(D, R)$ . . . 30
3.3.1 Estimation for a fixed (D, R) . . . 31

4 Filters and Diagnostics **36**
4.1 Univariate Light Curve Filters . . . 36
4.1.1 The Three Sigma Clipping Filter . . . 37
4.1.2 The EW filter . . . 38
4.1.3 Comparison . . . 38
4.1.4 Autocorrelation . . . 39
4.1.5 Adjustments To the Three-Sigma Filter . . . 39
4.2 Filters Based On Other Stars . . . 39
4.2.1 Heuristic Regression Argument . . . 40
4.3 Diagnostics . . . 41
4.3.1 Chi Square Test . . . 41
4.3.2 Stationary Versions of the Chi-Square Test . . . 42

5 Combination **53**
5.1 Simulated Occultations with TAOS data . . . 53

6 Single Point Detection **59**
6.1 Power in terms of c . . . 60
6.2 TAOS data and implications about power . . . 62

7 Fisher's Method **74**
7.1 Optimal Properties of Fisher's Method . . . 74

8 Multi Point Theory **80**
8.1 Setup and Notation . . . 80
8.2 Optimal Detection . . . 80
8.2.1 The Poisson Model . . . 81
8.2.2 The Normal Model . . . 81
8.2.3 Computation and Distribution . . . 83
8.3 Sub-Optimal Methods . . . 84
8.3.1 Midpoint Reduction . . . 84
8.3.2 The Norm Test Statistic . . . 84
8.3.3 Linear Regression . . . 86
8.4 Sensitivity of Occultation Surveys . . . 88
8.4.1 The Norm Statistic and Survey Sensitivity . . . 89
8.5 Estimation of θ . . . 91

9 Multi Point Simulation **93**
9.1 A vulnerability of the *norm* statistic. . . . 93
9.2 Single Point VS Multi Point for TAOS . . . 94
9.3 Future Surveys: TAOS II . . . 94
9.3.1 Dim Stars . . . 96

9.3.2 Practical Issues and Considerations 97

10 Appendix **104**
10.1 Union Vs Exact 104
10.1.1 Distribution of the Minimum 104
10.1.2 Bounding the Relative Error 105
10.1.3 Computation of Intersections 105
10.1.4 Bounding Relative Error by g 106
10.1.5 Computation of g 106
10.2 The $LogSumExp()$ function 108
10.3 Importance Sampling the tails of LLR 109
10.3.1 Importance Sampling: The basic idea 109
10.3.2 Tail Probabilities for the LLR 109
10.3.3 Simulation for the Normal model 110

Bibliography **111**

Chapter 1

Background and Motivation

In this chapter we give some scientific background, explain the occultation method and the TAOS project, describe the type of signals that occultations produce, and outline the statistical issues that are addressed in the later chapters.

1.1 Scientific Background

1.1.1 Trans-Neptunian Objects

Trans-Neptunian Objects (TNO's) are objects in our solar system that orbit the sun and are on average further from the sun than Neptune; that is, further than 30AU [1]. TNO's are sometimes roughly divided into Classical and Resonant Kuiper Belt Objects (KBOs), Oort Cloud objects, and Scattered Disc Objects. Figure 1.1 shows some of these TNO's in our solar system.

Kuiper Belt objects are generally at distance 30-55 AU, have nearly circular orbits, and lie close to the plane of the solar system (the ecliptic plane). Neptune has negligible gravitational influence on many of these objects, but some objects are locked in synchronized motion with Neptune and are called resonant KBO's. For example, Pluto in figure 1.1 is a resonant KBO because it orbits the sun two times for every three Neptune orbits, known as 2:3 resonance. There are many objects in 2:3 resonance with Neptune and there are also many with other orbital ratios such as 1:2 or 3:5.

The Oort Cloud is a hypothesized cloud of comets at distances of tens of thousands of AU's that lies at the gravitational boundary of our solar system.

Scattered disc objects have higher eccentricities; they can overlap with the Kuiper belt at perihelion and be several times farther at aphelion. These objects can be further from the ecliptic plane and unstable due to the influence of Neptune. Sedna in figure 1.1 is sometimes classified as a scattered disc object, other times it is called a detached object, or inner Oort cloud object because it is too remote to be influenced by Neptune. Naming conventions for objects outside the classical Kuiper Belt but closer than the hypothesized

[1] An 'Astronomical Unit', AU, is the average distance from the earth to the sun, which is $1.5 * 10^{11}$ meters or 92.96 million miles

Oort Cloud are not widely agreed upon, but we will refer to them as *Inner Oort Cloud Objects*.

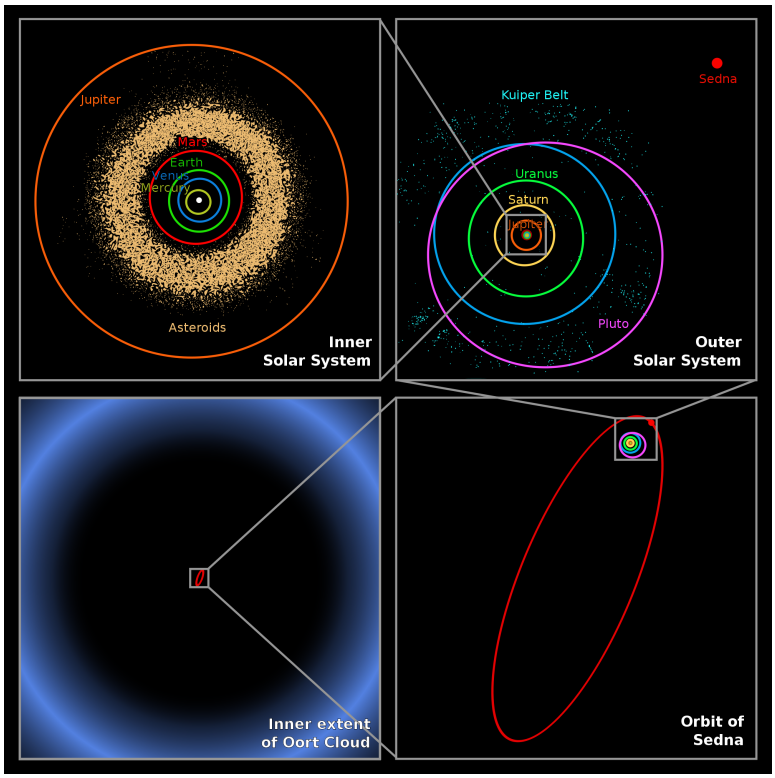

Figure 1.1: Diagram of some TNO objects. Neptune's orbit is represented by the blue path in right hand plots. Our primary interest is in the Kuiper Belt region in the upper right hand plot. Plot courtesy NASA/JPL-Caltech.

1.1.2 Population Quantities

In addition to learning about individual TNO's, there is also scientific interest in the distribution of objects in the outer solar system. The distribution of these objects contains information about the dynamics and evolution of our solar system; in particular the history of agglomeration and destructive collisions.

For objects with diameter greater than 100 km, observations are consistent with a power law density per square degree for the number of objects N with diameter D: $\frac{dN}{dD} \propto D^q$. Such specifications are sometimes written in cumulative form $N(D > c_1)$, or in terms of magnitude $N(R < c_2)$ [2].

In the past 6 years there have been many empirical findings about the population of smaller KBO's, and the general result is that there are far fewer small objects than would

[2]In Astronomy, magnitude (R) is a logarithmic measure of brightness in a certain filter band. Smaller magnitudes correspond to brighter objects. In fact, magnitude is what is observed and diameter is inferred based on distance to the object and the object's reflectivity (albedo), commonly assumed 4% for KBO's

be predicted by extending the power law to smaller diameters. TAOS is primarily interested in the population of KBO's, but has also done analysis for objects in the Inner Oort Cloud. Here I will primarily refer to the search for KBO's and abstractions of the process.

1.1.3 Direct Observations

Since the surface area of a sphere scales like r^2, the flux (photons collected) from a point source of light will decrease as r^2. When we see an object by observing reflected sunlight, and those reflected photons will also decrease like r^2 traveling from the KBO to earth, which implies [3] that flux decreases like r^4. In addition, TNO's are in motion relative to observation from the earth, so increasing exposure time is not an effective strategy. For these reasons, direct observation of small and distant TNO's will remain impossible for the foreseeable future.

The limits of direct observation have been pushed by [Bernstein et al., 2004] in a survey from the Hubble Space Telescope. They found three new objects, the faintest with magnitude 28.3, which corresponds to diameter 25 km given 4% albedo. By tracking the movement of objects across the images, the orbits could be determined, and the objects are depicted in figure 1.2.

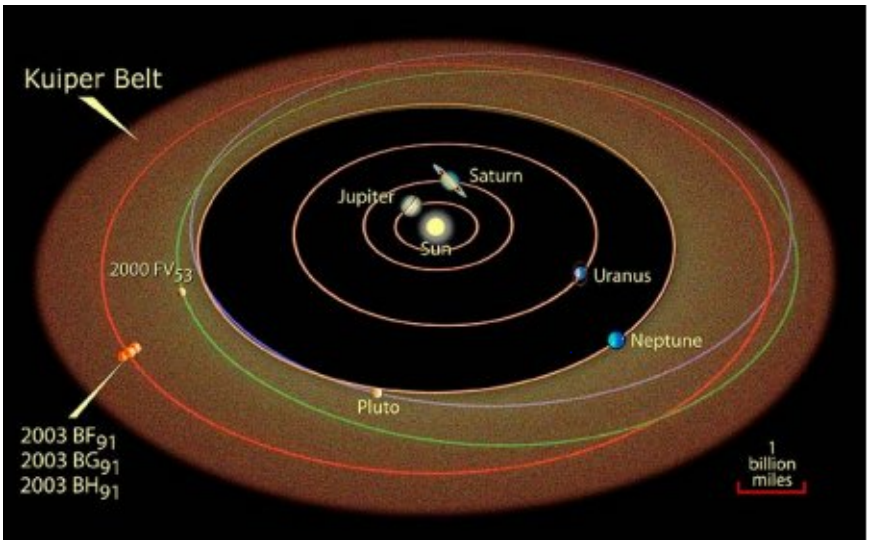

Figure 1.2: Artist illustration of the Kuiper Belt with three detections from 2003 [Bernstein et al., 2004]. The three detections are near each other because the observations were of fixed fields over a four day period, and estimates of orbital parameters gave distances in $40km - 43km$. The estimated eccentricities, however, were different for the objects and this is not depicted in the plot. Plot courtesy NASA.

[3] If r_1 is the distance from the sun to the KBO, and r_2 is the distance from the KBO to earth, then flux decreases like $r_1^2 * r_2^2$. If we assume the distance to the object is large compared to the distance between the sun and earth, then $r_1 \approx r_2$.

1.1.4 Indirect Observations: Occultations

The basic idea of the occultation technique is to observe a distant star and notice a change in its intensity when an object obstructs the line of sight. This is illustrated in figure 1.3.

When an occultation occurs, a diffraction pattern will be observed from earth, and a first order approximation is a momentary decrease in flux. For spherical objects, the diffraction pattern is known analytically, and is described in detail in [Nihei et al., 2007]. It is also known that the diffraction pattern of a small non-spherical object behaves like a spherical object of the same diameter.

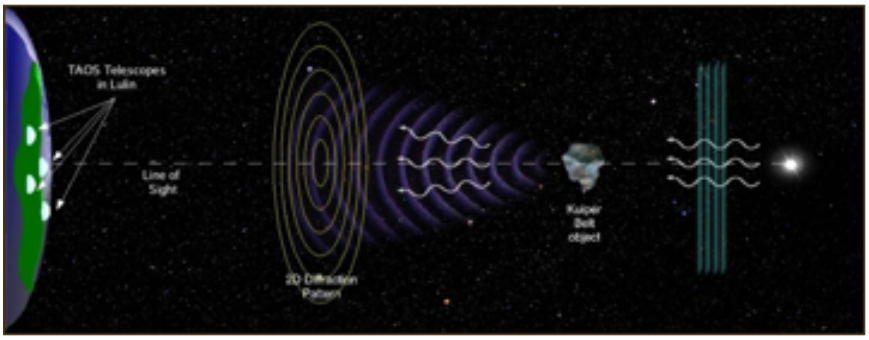

Figure 1.3: Illustration of an occultation and the resulting 2-D diffraction pattern. Plot courtesy of the TAOS collaboration.

In principle, if a diffraction pattern for a small or spherical object is observed at a high sampling rate and with high signal to noise ratio, inference about its size and distance could be made based on the analytic formulas. Unlike direct observation which follow the motion of objects, the occultation technique cannot give the orbit, but it can still aid in inference about the distribution of objects (population quantities).

1.2 The TAOS Project

The primary scientific goals of the TAOS project are to understand the distribution of small bodies (.5km - 10km) in the classical and extended KBO region, and to better understand the distribution of inner Oort Cloud objects. To do this, a system needed to be designed to monitor a large number of stars at a high sampling rate and to have a low false positive rate and a robust detection procedure.

The TAOS observations are taken from atop Lu-Lin Mountain in the Yu Shan area of Taiwan. Three telescope operations started in early 2005, and a fourth telescope came online in August 2008. Observation with multiple telescopes allows robust detection and stronger control of the false-positive rate by requiring simultaneous detection on all the telescopes. The telescopes have aperture of 50 cm and the field of view is approximately 3 square degrees, which allows observation of a large number of stars. A filter is used to capture light in the range 500nm - 700nm; a wide band is chosen to get more photons, implying a higher signal to noise, and the lower threshold was chosen to eliminate sky

background noise. The camera operates at 5 Hz, which allows detection of bodies in the .5km-10km regions. Details of the system are described in [Lehner et al., 2009].

As reported in [Bianco et al., 2010], for the 3.5 year period from Feburary 7, 2005 through August 2, 2008 the total exposure (star hours) was 500,339, which corresponds to about $9 * 10^9$ triplets in the three telescope data. Most stars observed are near the ecliptic and most opposition angles are small. Over 80% of the stars have magnitude in the range 12-14, and less than 3% are brighter than magnitude 10. For the dim stars (magnitude 12-14), signal to noise ratios are generally in the range 2-5, although 5% are greater than 10. For the stars with magnitures 10-12, signal to noise are generally around 10-25. More information on magnitudes and signal to noise ratios appears in the Chapter 6.

1.2.1 Photometry and Zipper Mode

The cameras essentially take a picture of a portion of the sky every fifth of a second, and each picture looks like the left hand side of figure 1.4. Photometry is basically the process of turning the series of pictures into a multivariate time series, where each time series is the series of calibrated flux measurements for each star. This problem is quite complex because there are background photons coming from the moon and other sources, and because changing atmospheric conditions can cause time dependent extinction and change the relative positions of the stars.

To get observations at 5 Hz from a normal CCD camera, TAOS uses a novel readout procedure called *zipper mode*. After 105ms, 76 rows are read out while leaving the shutter open, and the readout process takes 95ms. These 76 rows that are read out together are called a *row block*. Since the CCD has 2052 rows, it is not until $2052/76 = 27$ row blocks have been read out that all electrons from the first exposure have been removed from the CCD. Hence, in a row block image from zipper mode, the y-positions are stacked and the time stamps corresponding to a given star depends on its y-position in the stare mode image. Some portions of row block images are plotted in sequence in the right hand panel of 1.4. The circle in the right hand panel corresponds to the two circles in the left hand panel, and show how stars that are separated in space can become crowded in zipper mode images. The details of the custom zipper mode photometry are beyond our scope, but are explained in depth in [Zhang et al., 2009, Lehner et al., 2009, Zhang, 2009].

In addition to the challenges of zipper mode operation, the weather conditions at Lu-Lin can be bad [4]. The result is that TAOS photometry is extremely challenging and that resulting light curves need additional filtering, as will be discussed in Chapter 4.

1.3 Occultation Patterns and Parameterization

When an object occults a distant star relative to observation on earth, a two-dimensional diffraction pattern is sent in the direction of the line of sight from the star through the KBO, as shown in the left hand panel of figure 1.5 for a 3 km object at 42 AU. The one-dimensional diffraction profile observed from such a pattern depends on many

[4] Some time elapsed video of TAOS observing conditions can be found at http://www.youtube.com/user/hclin098

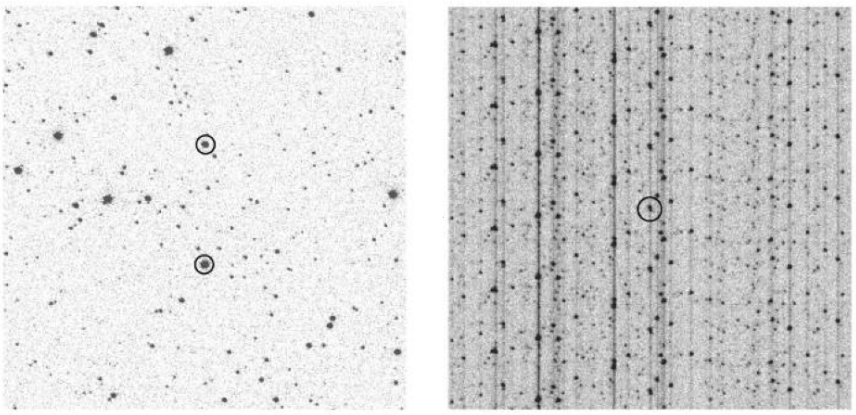

Figure 1.4: *Left Panel:* Subimage of a stare mode image. *Right Panel:* A series of zipper mode blocks, each of 76 rows, stacked on top of each other. Plot from [Zhang et al., 2009], courtesy of Z. Zhang

parameters, some come from the experimental setup and others remain unknown. We use the notation $I(t_i, \theta)$ to denote a **Diffraction Profile** for parameter θ as in the right hand side of figure 1.5. In this parameterization the baseline is 1, so $I(0, \theta) = .8$ would be interpreted as a 20% flux reduction at time $t = 0$.

The shape of the occultation depends on the size of and distance to the occulting object, the relative velocity of the occulting object, magnitude and spectral properties of the background star, the filter band used and other properties of the cameras, and the impact and timing parameters for the occultation.

The *impact parameter* is the minimal distance from our observing site to the line of sight of an occultation. As the earth rotates around the sun, if our telescope comes in the perfect line of sight of the observation, then the impact parameter is zero, and we observe along the red line of the left hand panel of figure 1.5, resulting in the red curve of the right hand panel of 1.5. If our closest point to the line of sight is 1 km, then we would observe along the blue line in the left hand panel, resulting in the blue curve in the right hand panel of 1.5. For large impact parameters, the observed diffraction profile is negligible. In [Nihei et al., 2007], it was found that $[0, H/2]$ was a reasonable interval for possible impact parameters, because the signal strength becomes small for larger values. Here, $H = H(D, \theta^*, d, \lambda)$ is the diameter of the first Airy ring, which depends on the size of the object, the angular size of the star, the distance to the object, and the median wavelength of the filter used on the telescope, respectively.

The *relative velocity* is the velocity (relative to earth) of the object perpendicular to the line of sight from earth. Assuming the earth and the KBO are in circular orbits, one can calculate the relative velocity as a function of the angle of observation, or *opposition angle* ϕ. If we look directly away from the sun then we have opposition angle zero. In Chapter 3 we derive the expression

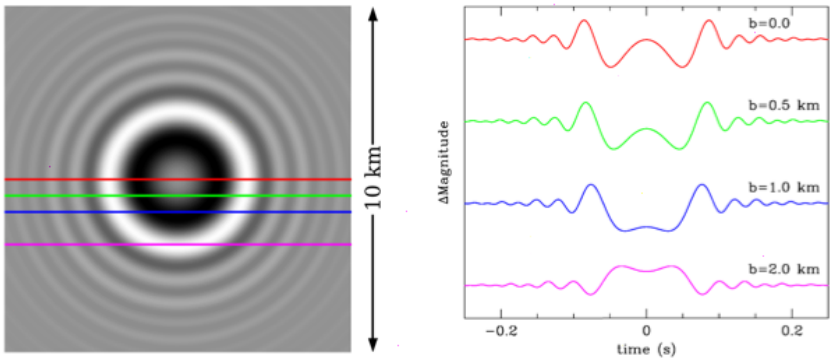

Figure 1.5: *Left Panel:* Two-Dimensional Diffraction pattern of a 3 km object at 42 AU. Horizontal lines correspond to different impact parameters. *Right Panel:* Diffraction Profiles produced by the impact parameters in left panel. Plot from [Bianco, 2009], courtesy of F. Bianco.

$$v^* = v_{earth} * \left[cos(\phi) - \sqrt{\frac{1}{\Delta_{au}}(1 - (\frac{1}{\Delta_{au}}sin(\phi))^2)}\right]$$

Where Δ_{au} is the distance to the kbo in AU's. Some calculus shows that this is increasing in distance once $\Delta_{au} > \sqrt{3}sin(\phi)$, so in our region of interest farther objects move faster, but not by a drastic amount. Also, we note that since Δ_{au} is relatively large, this velocity is approximated by $v_{earth} * cos(\phi)$, and we recall $v_{earth} \approx 30\frac{km}{sec}$.

The *timing parameter* refers to the fact that the CCD in TAOS has some down time; charge is being transferred for 95ms of every 200ms. The offset between the beginning of the transfer time and the time the telescope is closest to the line of sight is called the timing parameter, and it influences the pattern that will be seen.

In figure 1.6 we look at diffraction profiles for a magnitude 11 star with spectral type F0V by a 3km object with impact and timing parameters of zero. The figure illustrates differences in the profile seen by survey sampling rate, opposition angle, and distance to the occulting object. The 5Hz sampling rate imitated TAOS, and the 40Hz sampling rate illustrates a hypothetical survey where we assume there is negligible down time (hence no timing parameter). In figure 1.7 we look at the types of signals in the upper left hand part of figure 1.6, where we allow the impact and timing parameter to vary.

For a given lightcurve, we know or have estimates for all the parameters that influence θ except for timing parameter, impact parameter, size, and distance, which we denote (t, i, s, d). In principle, our parameter space could be

$$\Theta = (0, 95ms) \times (0, \infty) \times (0, \infty) \times (0, \infty)$$

However, we know that diffraction profiles for impact parameters greater than $H/2$ are weak, so we restrict the impact parameter to the interval $(0, H/2)$. Also, we do not anticipate objects bigger than the earth, or to use the technique outside the Oort Cloud, so

size and distance should not extend to infinity. A more reasonable parameter space for our purpose is

$$\Theta = (0, 95ms) \times (0, H/2) \times A$$

Where $A \subset (0, 10^4 km) \times (30AU, 10000AU)$ is not necessarily a square. Exactly where the subset A lies depends on parameters of the survey and the star. In [Zhang et al., 2008, Bianco et al., 2010], we consider $A \subset (.5km, 30km) \times (42AU)$, while noting that the diffraction profiles look similar for distances around $42AU$. In [Wang et al., 2009] we consider $A \subset (.5km, 30km) \times (100AU, 1000AU)$.

We will sometimes take a purely frequentist view of θ. Other times we will allow θ to have a prior over a finite or a continuous set of parameters. For the prior, it is reasonable to assume that objects at a fixed distance are uniformly distributed near the ecliptic, so the prior for impact and timing parameter should be uniform and independent of size and distance.

$$p(t, i, s, d) \propto p(s, d)$$

Finally, we mention that in our setup the null hypothesis of no occultation corresponds to $I(\theta) = \mathbf{1}$. Unfortunately, this does not correspond to a θ_0 in the interior of our parameter space for θ, as is often the case in hypothesis testing problems.

1.4 Models for Occultation Data

Since variable stars are excluded from being background stars in occultation studies, our observations under the null of no occultation are some noisy version of μ, the average number of photons per time point. Under the alternative θ, our observations should be some noisy version of $\mu * I(t_i, \theta)$.

If there was no atmosphere to worry about, and the cameras operated with perfect efficiency, then the arrival of photons would follow a Poisson process. That is, the number of photons that arrive during exposure i would follow:

$$y_i \sim \text{Poisson}(\mu I(t_i; \theta))$$

This Poisson model is an idealization, and the noise level of real data will always be higher. Even with data from the Hubble Space Telescope, [Schlichting et al., 2009], claim their data is 4% above the Poisson noise level. For ground based surveys, the noise level is much higher than the Poisson noise level.

A slightly more general model for a diffraction profile of length n is

$$y_t = \mu I(t; \theta) + e_t \qquad t = 1, ..., n \tag{1.1}$$

Where e_t is a noise term which could have changing variance and autocorrelation. If we make a normal approximation to the Poisson process above, this becomes

$$e_t \stackrel{d}{=} \sqrt{\mu I(t; \theta)} * \epsilon_t \quad \text{where} \quad \epsilon_t \sim \mathcal{N}(0, 1)$$

Since many photometry packages produce standard errors for the flux values they report, we could treat those as known constants and have

$$y_t = \mu I(t;\theta) + \sigma_t \epsilon_t \quad \text{where} \quad \epsilon_i \sim \mathcal{N}(0,1)$$

For TAOS data, the photometry does not produce reliable standard errors, so we also consider a model where σ_t is constant [6] during an occultation. We find it convenient to re parameterize this by

$$z_t = \frac{y_t - \mu}{\mu} \quad , \quad C(t_i;\theta) = I(t_i;\theta) - 1 \quad , \quad \lambda = \frac{\mu}{\sigma}$$

Giving

$$z_t = C(t;\theta) + \frac{1}{\lambda}\epsilon_t$$

These parametric models are studied in Chapter 6 for the single point case, and in Chapter 8 for the multi-point case. Models based on the normal distribution are unrealistic simplifications, and should be especially untrustworthy for calculating tail events. Our perspective is that such models can be useful in understanding the types of procedures that might perform well, but that significance levels should be calculated non-parametrically.

1.5 Statistical Questions

A number of statistical questions arise in addressing the scientific goals of TAOS, including

1. For multi-telescope data coming from models similar to that in equation 1.1, what is a good detection algorithm where the false positive rate can be calculated under reasonable assumptions?

2. What kinds of trends and anomalies remain in the light curves after the aperture photometry, and what processing steps are necessary so the data can be plausibly modeled according to the assumptions in (1)? Also, how can we test for violations that remain after our processing steps?

3. Since the atmosphere induces common signals on many stars during the same zipper run, can removing common signals improve signal to noise, and hence improve detection efficiency.

4. Given a detection pipeline and given 0 or $D > 0$ detections, what constraints are implied about the parameters of KBO populations?

5. What can be said in general about combining data from the K telescopes?

[6] If the errors were only slightly higher than the Poisson level, then this could be an overstatement of errors for deep occultations

6. What should be done if the data followed a parametric model, and what would that imply about TAOS and other occultation surveys?

7. How does the power of our non-parametric methods compare with optimal parametric procedures for real data and in the case where the parametric models are true?

For point (1), we find the *rank product* method to be an effective test statistic and prove that false positive rates can be calculated under the null hypothesis that individual time series are long and stationary [5], and that there is independence between series. The idea of using ranks in occultation detection was pioneered in [Liang, 2001] and [Liang et al., 2004], where they propose the *maximum rank* method. Details of the *rank product* method appear in Chapter 2, and a comparison of the methods is covered in Chapter 5.

Chapter 4, addresses point (2) above, Chapter 3 addresses point (3), and Chapters 8 and 9 address point (6). Chapter 7 addresses point (5) by deriving an optimality property of Fisher's Method, which is the motivation for the rank product method. Chapter 5 addresses point (5) from the point of view of TAOS data.

Regarding (3), we explored this issue but were not very successful in increasing the power of detection. This is explained in Chapter 4, and a heuristic argument based on the normal distribution suggesting there is not much to gain is given in Chapter 6. Issue (7) is also addressed in Chapter 6.

Statisticians and Astronomers considering detection based on independent sources may be interested in the material on the rank product distribution, particularly in the weak conditions under which it is valid. TAOS scientists considering filtering based on other stars should find the material in Chapters 4 and 6 of interest, and the material in Chapter 9 should be of interest in the search for distant objects, and also for TAOS II. Chapter 7 is probably of greater interest to Statisticians than Astronomers.

[5] The series must also be ergodic in mean, as explained in Chapter 2

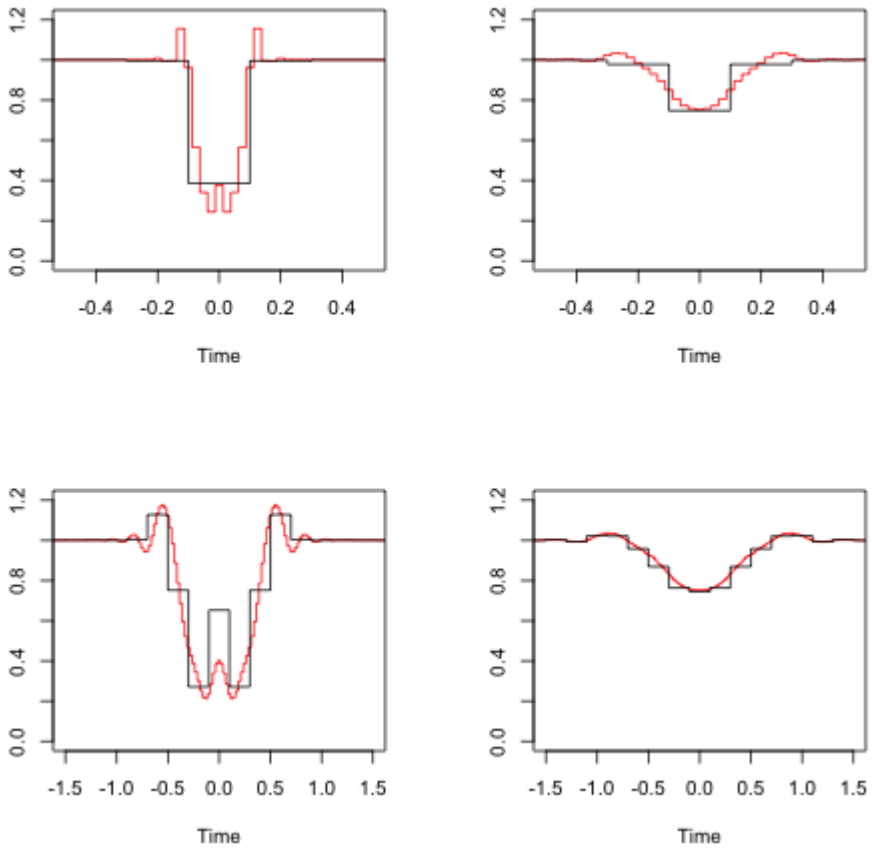

Figure 1.6: Occultation profiles of magnitude 11 star with spectral type F0V by a 3km object with central impact and timing parameters. **The red curves** indicate a survey running at 40 Hz, whereas black represents 5 Hz. **The top row** indicates observations taken at opposition, whereas the bottom are at opposition angle 70°. **The left column** indicate the occultation was at 43AU, whereas the right column had occultations at 300AU.

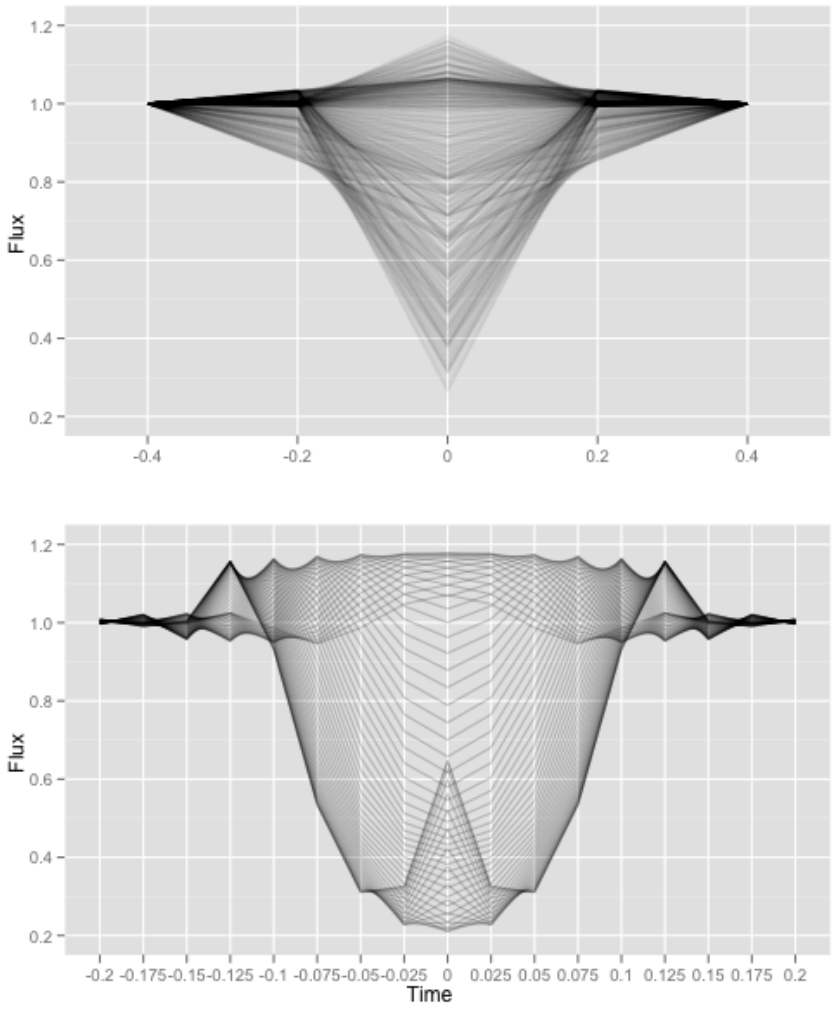

Figure 1.7: Occultation profiles of magnitude 11 star with spectral type F0V by a 3km object for 5Hz observations (top) and 40Hz observations (bottom). In the top, 60 equally spaced impact parameters and 17 equally spaced timing parameters are displayed. In the bottom, there is no timing parameter, but 60 equally spaced impact parameters are displayed.

Chapter 2

The Rank Product Method

2.1 Motivation for the Method

To an approximation, searching for small KBO's in TAOS data is like searching for a single point flux reduction across multiple telescopes. Since hypothesis tests are performed at each time point for every star observed, a high significance threshold must be set to keep the false positives at a reasonable level. In particular, we are interested in thresholds in the range $10^{-11} - 10^{-10}$.

While the bulk of the distribution of light curves follow a bell shaped curve, the tails are far wider than the Normal distribution, so tests based on Poisson or Normal statistics would not be believable at levels like 10^{-10}. The rank product method works if each time series is long and stationary [1], and the K time series are independent, so no assumptions need to be made on the marginal distribution of the data.

We note that the validity of the method under stationarity allows extensions of the method beyond single point flux reduction. If the flux measurements are stationary, and we take a running function of the data (like a moving average or a convolution), then the resulting series will remain stationary and the rank product can be used. In searching for longer events we would pick a function that we expect to respond well to such events.

2.2 The Rank Product Method

Assume we have K time series of length N, where each column represents a time series.

$$X = \begin{pmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,K} \\ x_{2,1} & x_{2,2} & \dots & x_{2,K} \\ \vdots & \vdots & \vdots & \vdots \\ x_{N,1} & x_{N,2} & \dots & x_{N,K} \end{pmatrix}$$

[1] We also need ergodicity in mean; so there can't be extremely long range dependence and we need existence of the first moment.

Now suppose that each time series observation is replaced by its rank within that time series, giving the matrix:

$$R = \begin{pmatrix} r_{1,1} & r_{1,2} & ... & r_{1,K} \\ r_{2,1} & r_{2,2} & ... & r_{2,K} \\ \vdots & \vdots & \vdots & \vdots \\ r_{N,1} & r_{N,2} & ... & r_{N,K} \end{pmatrix}$$

Now suppose that at each time point, we look at the product of the ranks:

$$t_i = \prod_{j=1}^{K} r_{i,j}$$

Which gives us

$$T = (t_1, t_2, ..., t_N)'$$

A low value of t_j means the ranks were low across all the time series at time j, showing evidence of a simultaneous drop at time j. The final step is to calculate the p-values corresponding to the test statistic:

$$X_{N\times K} \longrightarrow R_{N\times K} \longrightarrow T_{N\times 1} \longrightarrow P_{N\times 1} = (p_1, ..., p_N)'$$

2.3 Calculation of P-Values

We assume that the original data matrix X is such that at each time point i, the k-tuple $(r_{i,1}, ..., r_{i,K})$ is uniform on $\{1, ..., N\}^K$.

If for each k we had $x_{1,k}, ..., x_{N,k} \overset{iid}{\sim} F_k$, for some continuous F_k and we had independence between the time series, then this would hold. In fact, it holds as long as the individual series are long and stationary, as will be shown below.

2.3.1 A Simple Example

Suppose that K=3, and $t_j = 3$, then the combinations of ranks that could lead to something this small or smaller are:

$$(1,1,1),\ (2,1,1),\ (1,2,1),\ (1,1,2),\ (3,1,1),\ (1,3,1),\ (1,1,3)$$

There are N^3 possible combinations, and 7 that produce values ≤ 3, so $p_j = 7/N^3$.

2.4 Continuous Approximation: Fisher's Method

Fisher's Method of Combining Independent Tests of Significance is to use the product of the p-values as the test statistic. This method assumes that the underlying data is continuous, so that the p values are uniform on (0,1).

Assume U is uniform on $(0,1)$ then we have

$$-2\ln(U) \stackrel{d}{=} exp(1/2) = \chi^2_2$$

So if $p_1, ..., p_K$ are independent under the null, then

$$-2\ln(\prod_{i=1}^{K} p_i) \stackrel{d}{=} \chi^2_{2K}$$

It should be noted that in TAOS publications this is written in terms of the Gamma distribution and that the Chi-Squared Distribution is a special case of the Gamma distribution.

With single telescope data, one could give observation x_j a p value of r_j/N. Here the p-values are uniform on the discrete set $\{1/N, 2/N, ..., 1\}$. The product of p-values is proportional to the rank product: $N^{-K} * t_i$.

2.5 Quality of the Continuous Approximation

The exact p-values are given by

$$\mathbb{P}(t_i \leq c) = \frac{1}{N^K} \sum_{i=1}^{c} g(i, K, N)$$

Where

$$g(x, K, N) = \#\{(a_1, ..., a_K) \in \{1, ..., N\}^K : \prod_{k=1}^{K} a_k = x\}$$

The continuous approximation is

$$\begin{aligned} \mathbb{P}(t_i \leq c) &= \mathbb{P}(-2\ln(\frac{t_i}{N^K}) \geq -2\ln(\frac{c}{N^K})) && (2.1) \\ &\approx \mathbb{P}(\chi^2_{2K} \geq -2\ln(\frac{c}{N^K})) && (2.2) \\ &= pchisq(-2\ln(\frac{c}{N^K}), 2K, lower.tail = F) && (2.3) \end{aligned}$$

Where the last line is the corresponding R code. Figure 2.1 shows the exact pvalue and continuous approximation for three series of length 27000. Clearly the continuous approximation understates significance in the tails.

2.6 Exact Calculation of Tail Probabilities

To calculate the exact p-values, one needs to know $g(i, K, N)$. One method to do this is to run K nested for loops from 1 to N, look at the resulting products and keep tally of how many times you get each product.

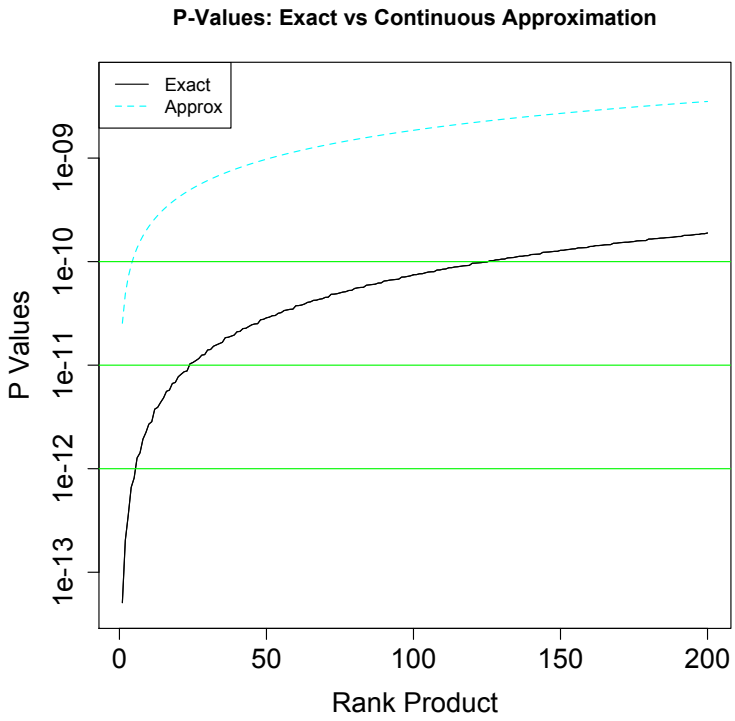

Figure 2.1: Exact P-Values and Continuous Approximation for three telescope data and series length 27000. The green lines indicate levels $10^{-10}, 10^{-11}, 10^{-12}$.

The for loop method is fast and practical, but we also consider another method for calculating $g(i, K, N)$ for $i \leq N$ that gives us a better understanding of the nature of the jumps in the distribution of p-values. We clearly have $g(x, K, N) = 1$ for $x = 1$. For $1 < x \leq N$, we have

Lemma 2.6.1. *Consider the prime decomposition of* $1 < x \leq N$ *where the* p*'s are unique primes and* d *is their degree so that:*

$$x = p_1^{d_1} * p_2^{d_2} * ... * p_m^{d_m}$$

Then

$$g(x, K, N) = \prod_{i=1}^{m} \binom{d_i + k - 1}{k - 1} = g(x, K)$$

The last equality is a definition of $g(x, K)$ where we have dropped the subscript N to note that this quantity doesn't depend on N as long as $x \leq N$.

Proof. Suppose $A_1 * A_2 * ... * A_k = x$ and take prime decompositions of each number:

$$\begin{aligned} A_1 &= p_1^{d_{1,1}} * p_2^{d_{1,2}} * ... * p_m^{d_{1,m}} \\ A_2 &= p_1^{d_{2,1}} * p_2^{d_{2,2}} * ... * p_m^{d_{2,m}} \\ &\vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ A_k &= p_1^{d_{k,1}} * p_2^{d_{k,2}} * ... * p_m^{d_{k,m}} \end{aligned}$$

$$\text{Where} \quad \sum_{i=1}^{k} d_{i,j} = d_j \quad \forall j$$

Hence,

$$g(x,k,N) = \prod_{i=1}^{m} S(d_i;k)$$

Where

$$S(d;k) = \#\{(a_1,...,a_k) \in \{0,..,d\}^k s.t. \sum_{i=1}^{k} a_i = d\}$$

The example (4,0,1,5) is a case of d = 10, k = 4, and can be illustrated by the picture

$$**** || * | ****$$

Since there are 4 stars to the left of the first |, then 0 between the next two, one between the next two and five after the last one. So the number of possible 4-tuples is the number of ways to choose 3 bar locations in the 13 possibilities. This gives

$$S(d;k) = \binom{d+k-1}{k-1}$$

□

2.6.1 Examples

Since $6 = 2*3$ and $22 = 11*2$ are both the product of two primes to the first power, they have the same number of ways to get that product, and in the 4 telescope case it is $(\binom{1+4-1}{4-1})^2 = 16$. For a more complex example, the number of ways to get $2520 = 2^3*3^2*5*7$ or $8505 = 3^2 * 3^3 * 5 * 7$ in the 4 telescope case is equal to $(\binom{3+4-1}{4-1})(\binom{2+4-1}{4-1})(\binom{1+4-1}{4-1})^2 = 20 * 10 * 4^2 = 3200$.

2.6.2 Important Cutoffs

Given only two telescopes, a rank of $(1,1)$ in series of length 27000 would give p value $1/27000^2 = 7.3 * 10^{-8}$. Given that we want significance levels like 10^{-10}, we see that detections are impossible using the rank product method with two telescope data. For three and four telescope data detections are possible, and the thresholds for various levels are given below.

Level	3 Telescopes	4 Telescopes
10^{-10}	125	142687
10^{-11}	23	21959
10^{-12}	5	3671
10^{-13}	1	663

2.7 Test at time points or over the whole series ?

Since occultations are such rare events, we can safely ignore the possibility that two occur in one light curve. Also, if an occultation by a large object causes the rank product to pass the threshold for several consecutive time points, then it would only get counted as one event. Because of these facts, it makes sense then to test a hypothesis over entire light curves rather than point by point. The relevant test statistic is the MINIMUM rank product over the entire light curve.

If we tested point by point at level α in a series of length N, then the contribution of the light curve triple to the false positive rate would be $N * \alpha$, so testing point by point is like using the union bound to approximate the distribution of the minimum.

Let r_i^k be the rank at time i on the k'th telescope. Let the rank product be $R_i = \prod_{k=1}^{K} r_i^k$, and let $X = \min_i R_i$. Which threshold should we base our detections algorithm on ?

$$\mathbb{P}(R_i \leq c_1) \leq \alpha \qquad \mathbb{P}(X \leq c_2) \leq N * \alpha$$

The answer is that it does not matter. For most choices of small α, we will have $c_1 = c_2$. In the case where underlying data is iid (hence the ranks are exchangeable), it is possible to calculate the distribution of X. It turns out that the relative error in using the union bound is less than 10^{-6}, or 1000^{th} of 1%, in the region we care about (see figure 2.2). Details of the computation appear in Appendix section 10.1.

There are jumps in the distribution due to discreetness of the rank product. The smallest jump occurs where the rank product is prime, where the CDF changes by $\frac{3}{N^3}$. At level 10^{-m}, this is a relative jump of $1.7 * 10^{m-14}$. Hence the smallest jump in the CDF due to discreteness is at least 50 - 1000 times bigger than the relative error in the union bound in our region of interest. This implies that we will almost always have $c_1 = c_2$.

2.8 Weaker Assumptions for the Rank Product Distribution

We know that if the time series of observations are i.i.d. and we have independence between series, then the rank product distribution is valid, but it is also valid under weaker assumptions. Our question is: for what joint distributions $(X_1, ..., X_N) \sim F$ will the distribution of the ranks at each time point $1 \leq i \leq N$ be uniform?

We need more than a common marginal distribution. For example, X_1 would be very unlikely to have rank 2 if (X_1, X_2, X_3) followed

$$X \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & .99 \\ 0 & .99 & 1 \end{pmatrix}\right)$$

We need more than stationarity, as is shown in the left hand column of figure 2.3. There we take a white noise series of length 7, take a five point moving average, then throw away the two endpoints on each side leaving a 3 point stationary series. In this case, extreme ranks are more likely at the end of the series because they have less competition. The right hand panel shows that the effect disappears when the MA(5) series is of length 1000 rather than length 3. This suggests that if the time scale of autocorrelation of a stationary series is small compared to the length of the series, then the ranks will be approximately uniform; this is proven below.

2.8.1 Exact Results

A sufficient condition for uniform ranks across a random vector is:

Definition 2.8.1. *We say $x_1, ..., x_N$ is Strongly Stationary on the Circle (SSC) if for all $1 \leq i \leq j \leq N$ and all m we have*

$$(x_i, ..., x_j) \stackrel{d}{=} (x_{[i+m]_N}, ..., x_{[j+m]_N})$$

Where $[k]_N$ is the remainder when dividing k by N.

We note that: i.i.d. implies *Exchangeable* implies *SSC*.

Lemma 2.8.2. *If each time series is SSC and the series are independent of each other, then at each time point j, the rank-tuple*

$$(r_{1,j}, r_{2,j}, ..., r_{K,j})$$

is uniform on

$$\{1, 2, ..., N_p\}^K$$

Proof. Suppose we replace $x_1, ..., x_N$ with its ranks, $r_1, ..., r_N$.

If the series is *SSC*, consider the functions f that takes a list of N numbers and returns the rank of the first one. Using the definition with $i = 1, j = N$, this gives

$$f(x_1, ..., x_N) \stackrel{d}{=} f(x_{[1+m]_N}, ..., x_{[N+m]_N})$$

Applying this for $m = 1, 2, ..., N$ gives us:

$$r_1 \stackrel{d}{=} r_2 \stackrel{d}{=} ... \stackrel{d}{=} r_N$$

Let $\mathbb{P}(r_t = k) = \mathbb{E}(\mathbf{1}_{r_t=k}) = c_k$, where c_k does not depend on t because of the equality in distribution. This implies that $\mathbb{E}(\sum_{t=1}^{N} \mathbf{1}_{r_t=k}) = N * c_k$. Since rank k must appear exactly once in the time series, $\sum_{t=1}^{N} \mathbf{1}_{r_t=k} = 1$, so we must have $c_k = \frac{1}{N}$. Hence, at each time point t, r_t is uniform on $\{1, 2, ..., N\}$.

Next, using independence of the series we get the desired result.

□

2.8.2 Asymptotic Results

For time series arising from flux counts with serial dependance, it is unlikely that the stationary structure extends circularly at the endpoints. However, if the series are long, and the dependance is not too strong, it seems reasonable to expect that the results should be approximately the same as those for the *SSC* series.

We assume there exists a discrete-time, stationary, time series on $(-\infty, \infty)$, and that we observe the subset

$$x_1, ..., x_n$$

We argue that if n is large and the dependance is not too strong, and if we replace the series with its ranks, then at each time point, the rank is approximately uniform on $1, ..., n$.

Comparison of MA type process to SSC process

One type of stationary process is those which are a rolling function of another process. If the window is small compared to the length of the series, then the series will still be uniform.

Lemma 2.8.3. *Suppose we replace an i.i.d. or exchangeable, time series $x_1, ..., x_{N+2k}$ by $y_t = f(x_{t-k}, ..., x_t, ..., x_{t+k})$, using the circular endpoint rule. The resulting series of y_t is of length $N + 2k$, but we write it in terms of endpoints and a center part:*

$$e_1, ..., e_k, y_1, y_2, ..., y_N, e_{k+1}, ...e_{2k}$$

and consider all elements to be the 'long series', and the middle part to be the 'short series'. We let r_i^S be the rank of y_i in the short series, and r_i^L be the rank of y_i in the long series. Then as $N \to \infty$ and $\frac{k}{N} \to 0$ we have

$$\frac{r_i^S}{N} \xrightarrow{d} \text{Unif}(0, 1)$$

Proof. Since the long series is SSC, we know that r_i^L is uniform on $\{1, ..., N+2k\}$. We must have:

$$r_i^L - 2 * k \leq r_i^S \leq r_i^L$$

Which implies

$$\frac{r_i^L}{N+2k} - \frac{2k}{N+2k} \leq \frac{r_i^S}{N}\frac{N}{N+2k} \leq \frac{r_i^L}{N+2k}$$

Letting U_x be uniform on $\{\frac{1}{x}, \frac{2}{x}, ..., 1\}$, and $\rho = \frac{k}{N}$, this implies

$$(1+2\rho)U_{N+2k} - 2\rho \leq \frac{r_i^S}{N} \leq (1+2\rho)U_{N+2k}$$

So if $N \to \infty$ and $\frac{k}{N} \to 0$ then we have

$$\frac{r_i^S}{N} \xrightarrow{d} \text{Unif}(0,1)$$

□

Results for General Stationary Process

For stationary process, *Ergodicity in Mean* is sufficient to give uniform ranks in the limit.

Definition 2.8.4. *A stationary time series* $\{X_t : t = 0, \pm 1, \pm 2, ...\}$ *is Ergodic in Mean if for every* f *with* $\mathbb{E}f(X_t) < \infty$ *we have*

$$\frac{1}{N}\sum_{i=1}^{N} f(X_i) \xrightarrow{a.s.} \mathbb{E}f(X_t)$$

We start by establishing some lemmas:

Lemma 2.8.5. *Glivenko-Cantelli Lemma*
If a series is Ergodic in mean, then Glivenko-Canteli holds, namely

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \xrightarrow{a.s.} 0$$

Proof. The proof is identical to the i.i.d. case, which uses the law of large numbers on indicator functions while being careful about jumps in F (See Theorem 1.7.4 in [Durrett, 2005]).
□

Also, a standard $\epsilon - \delta$ argument that appears in many probability texts gives:

Lemma 2.8.6. *Converging Together Lemma*
If $X \xrightarrow{d} Z$ *and* $|X - Y| \xrightarrow{\mathbb{P}} 0$*, then* $Y \xrightarrow{d} Z$

Theorem 2.8.7. *Assume the stationary time series of continuous random variables* $\{X_t : t = 0, \pm 1, \pm 2, ...\}$ *is Ergodic in Mean, and let* r_i^N *be the rank of* X_i *in* $\{X_1, ..., X_N\}$ *where* $i < N$*. Then for any sequence* $i(N)$*, as* $N \to \infty$ *we have*

$$\frac{r_{i(N)}^N}{N} \xrightarrow{d} U$$

Proof. Letting F_n be the ecdf of $X_1, ..., X_n$, we know from Glivenko-Cantelli that

$$D_n \equiv \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \xrightarrow{a.s.} 0$$

Moving to the probability space of infinite sequences and letting $F_n(\omega)$ be the empirical distribution function at time n based on sample element ω, we introduce the random variable C_n by

$$\begin{aligned} D_n(\omega) &= \sup_{x\in\mathbb{R}} |F_n(\omega)(x) - F(x)| \\ &\geq |F_n(\omega)(X_{i(n)}(\omega)) - F(X_{i(n)}(\omega))| \\ &= |r^n_{i(n)}(\omega)/n - F(X_{i(n)}(\omega))| \\ &\equiv C_n(\omega) \end{aligned}$$

Since $D_n \xrightarrow{a.s.} 0$, we must have $C_n \xrightarrow{a.s.} 0$, which implies $C_n \xrightarrow{\mathbb{P}} 0$

However, the random variable $F(X_{i(n)})$ always has the same distribution by stationarity, and this distribution is uniform because F is continuous. Using the converging together lemma, we get

$$\frac{r^n_{i(n)}}{n} \xrightarrow{d} U$$

□

Mixing

One condition to assure *Ergodicity in Mean* is *α-mixing* (see [Ibragimov and Linnik, 1971])

Definition 2.8.8. *Given a strictly stationary time series* $\{X_t : t = 0, \pm1, \pm2, ...\}$*, and letting* $\mathcal{F}^b_a = \sigma(X_a, ..., X_b)$ *be the sigma field generated by* $X_a, ..., X_b$*, we say the time series is α-mixing if*

$$\alpha(n) = sup_{A,B}\{|\mathbb{P}(A)\mathbb{P}(B) - \mathbb{P}(AB)| : A \in \mathcal{F}^0_{-\infty}, B \in F^\infty_n\} \longrightarrow 0$$

In English, this says events separated by sufficient time are approximately independent.

As an example, suppose we replace an i.i.d. time series $..., x_{-n}, ..., x_0, ..., x_n, ...$ by $y_t = f(x_{t-k}, ..., x_t, ..., x_{t+k})$. Then for all $m > 2k$ we have $\alpha(m) = 0$, so the series is *α-mixing.*

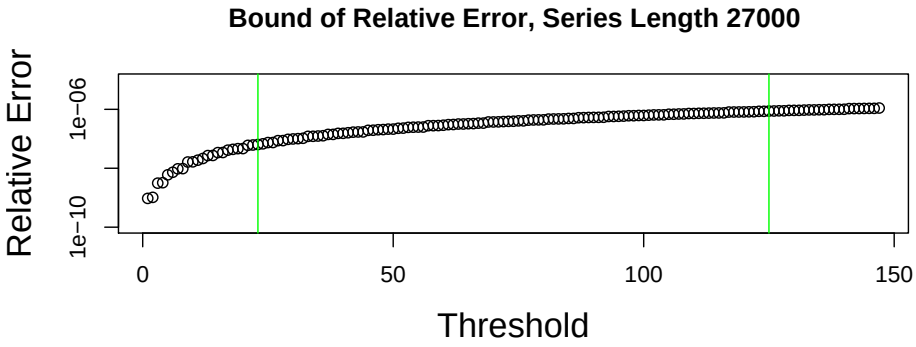

Figure 2.2: Bound on relative error of the union bound. The vertical lines correspond to the thresholds 10^{-10} and 10^{-11}

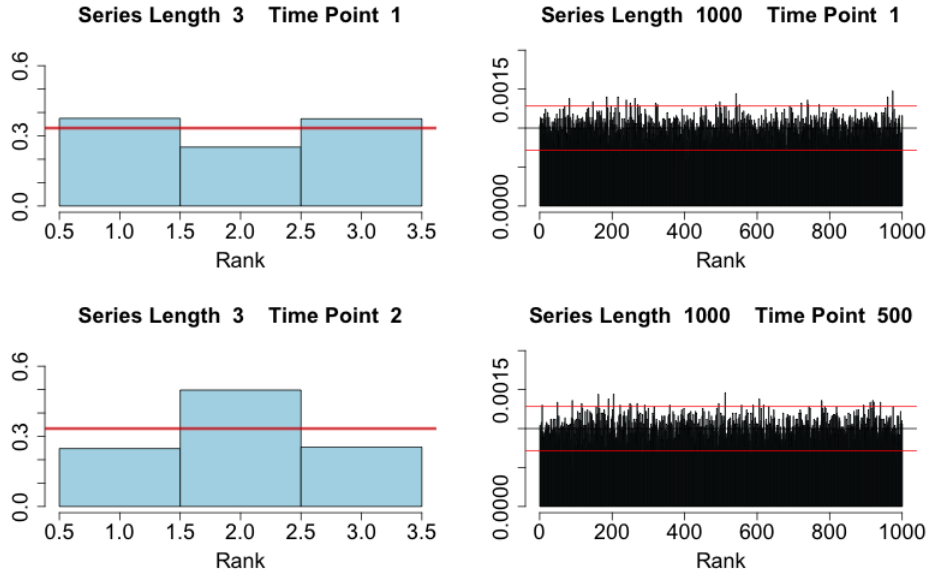

Figure 2.3: *Left Column:* We generate a length three $MA(5)$ process 5000 times, and look at a histogram ranks for a time points 1 (top), and 2 (bottom). The red lines correspond to 95% confidence intervals under the null that the ranks are uniform. *Right Column:* Same as left hand column but with a length 1000 $MA(5)$ process.

Chapter 3

Efficiency

In this chapter we show how the expected number of detections in an occultation survey is a function of the efficiency of our detection procedure, the population of TNO's, and the coverage of the survey. We also describe how the efficiency can be estimated, along with its standard errors. Given data from an occultation survey, we know coverage and number of detections, and we estimate efficiency, so inference about the population can be made.

3.1 Expected Number of Detections

Suppose our survey is made up of L light curves, or light curve triples/quadruples in the case of TAOS. We assume the survey operates at h Hz, and that light curve i has n_i points. We transform a light curve (or light curve triple) into a time series of test statistics $T_1, ..., T_{n_i}$ where extreme values of T_j indicate presence of the signal at time point j. A 'detection' is said to occur if T_j exceeds some threshold [1]. For the moment, we assume we know the distribution of T_j, and the level of our test is α.

Definition 3.1.1.

$$p_i(D, R) = \mathbb{P}(\textit{Detection of (D,R) in light curve } i)$$

More specifically, $p_i(D, R)$ is the probability that an occultation by a TNO at distance R and with diameter D in light curve i will be detected by our procedure. This probability has integrated out over uniform distributions on the time within the light curve where the occultation occurred, and the impact and timing parameters.

Definition 3.1.2.

$$M(D, R) = \textit{The density of TNO's by size and distance per square degree}$$

We implicitly assume we are near the ecliptic. Ecliptic latitude could be another variable in the density.

[1] In TAOS for three telescope data of length 27000 with level 10^{-10}, the correspond procedure is to filter, then rank the time series, then detect if the rank product is less than 125.

Definition 3.1.3. *Weighting Factor*

$$f_i(D,R) = \frac{E * v^* * H}{R^2}$$

Is the weighting factor for light curve i for at object at distance R with diameter D. Here $E = n_i * h$ is the exposure time, v^* is the relative velocity perpendicular to the line of sight, and

$$H \approx \left((2\sqrt{3}F)^{\frac{3}{2}} + D^{\frac{3}{2}} \right)^{\frac{2}{3}} + \theta_* R$$

is the event cross section as explained in [Nihei et al., 2007]. Here θ_* is the angular diameter of the star and F is the Fresnel Scale: $F = \sqrt{\frac{\lambda R}{2}}$, where λ is the median wavelength used in the survey (for TAOS, this is $\approx 600nm$).

Then for a survey, we have

Lemma 3.1.4.

$$\mathbb{E}\big[\#\ of\ detections\ in\ survey\big] = \int M(D,R)\Omega_e(D,R)dD + \alpha * T_{tot}$$

Where

$$\Omega_e(D,R) = \sum_{l=1}^{L} (p_l(D,R) * f_l(D,R))$$

*Is the **Effective Solid Angle** of the survey, and*

$$T_{tot} = \sum_{i=1}^{L} n_i$$

We note that one could perform many different test, some working better for large objects rather than small, and lemma 3.1.4 will remain valid as long as the level α refers to the chance of a false positive under **any** of the many tests performed.

3.1.1 Simple Illustration of Efficiency

The cartoon in figure 3.1 illustrates observations from the Earth of a distant star on one night, where the shell is a portion of the Kuiper belt that may contain KBO's.

We assume there are exactly M 3km objects, uniformly distributed in the shell, and no other objects. We let N be the number of objects that pass between the telescope and the star, and let p be the probability of detecting such an object in this light curve. Then

$$
\begin{aligned}
\mathbb{E}\big[\#\text{ of detections}\big] &= \mathbb{E}\big[\#\text{ of real detections}\big] + \mathbb{E}\big[\#\text{ of false positives}\big] \\
&= \mathbb{E}\big[\#\text{ of real detections}\big] + T\alpha \\
&= \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{N} \mathbf{1}_{\text{Object i is detected}} | N\right]\right] + T\alpha \\
&= \mathbb{E}\left[N * \mathbb{E}\left[\mathbf{1}_{\text{Object 1 is detected}} | N\right]\right] + T\alpha \\
&= p\mathbb{E}\Big[N\Big] + T\alpha \\
&= p * M * \frac{\text{Volume of Shell Seen}}{\text{Volume of Shell}} + T\alpha \\
&= p * M * f + T\alpha
\end{aligned}
$$

Where we have defined f as the *volume fraction* appearing in the second to last line.

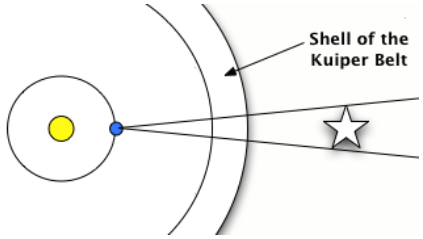

Figure 3.1: Not to scale diagram of volume observed

3.1.2 Proper Evaluation of f

As the shell width gets small in figure 3.1, our fraction problem turns two dimensional, as we are looking at a surface. In addition, we are not looking at the entire sphere, but just the sliver near the ecliptic plane. Based on the uniformity assumption, it is better to think of M as the number of objects *per square degree* near the ecliptic and then we have,

$$\mathbb{E}(N) = M * f \equiv M * \textit{square degrees covered}$$

The computation above of the volume fraction term f did not take into account the relative motion of objects. The portion of the sky we can detect perpendicular to our line of sight is the product of

$$\begin{aligned} E &= \text{Length of Run (in seconds)} \\ v^* &= \text{Relative Velocity } \perp \text{ sight (in meters/second)} \\ H &= \text{Diameter of 1st Airy Ring } ^2 \text{ (in meters)} \end{aligned}$$

The distance vectors $E * v^*$ and H are perpendicular, and each can be converted to degrees through the small angle formula:

$$f = \frac{E * v^* * H}{R^2} = \frac{E * v^*}{R} * \frac{H}{R} = \text{Square Degrees Covered}$$

3.1.3 Relative Velocity and Opposition Angle Geometry

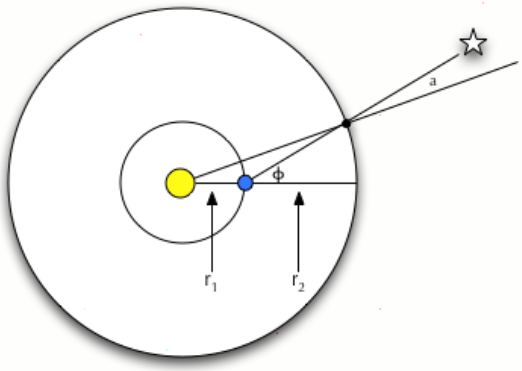

Figure 3.2: Diagram of opposition geometry. The central circle represents the sun, the next circle represents the orbit of the earth, and the outer circle represents the orbit of a KBO. If viewed above the north pole, motion is counter clockwise.

In figure 3.2, the angle ϕ is called the opposition angle, it points from earth E toward a distant star. An opposition angle of zero means we are looking directly away from the sun. The opposition angle and distance determine v^* for objects in circular motion:

Lemma 3.1.5. *For the earth and an object at Δ_{au} AU in circular orbits, the relative velocity perpendicular to the line of sight of earth is given by*

$$v^* = v_{earth} * \left[cos(\phi) - \sqrt{\frac{1}{\Delta_{au}}(1 - (\frac{1}{\Delta_{au}} sin(\phi))^2)} \right]$$

Proof. We assume that the KBO and earth are both in counter-clockwise circular orbits. The earth will have velocity v_{earth} tangential to the inner circle, while the KBO (K in the diagram) will have velocity v_{kbo} tangential to the outer circle. The component of the KBO's velocity perpendicular to the line of sight is $v_{kbo} * cos(a)$. The component of the

earth's velocity perpendicular to the line of sight is $v_{earth} * cos(\phi)$, so the relative velocity perpendicular to the line of sight is $v^* = v_{earth} * cos(\phi) - v_{kbo} * cos(a)$.

To simplify this, we first use the law of sines to note that

$$\frac{sin(180-\phi)}{r_1+r_2} = \frac{sin(a)}{r1} \qquad \text{hence} \qquad sin(a) = -\frac{r_1}{r_1+r_2} sin(\phi)$$

This implies $cos(a) = \sqrt{1-(\frac{r_1}{r_1+r_2} sin(\phi))^2}$

Also, Newton's laws of motion give $v_{kbo} = v_{earth}\sqrt{\frac{r_1}{r_1+r_2}}$

Putting it together,

$$\begin{aligned} v^* &= v_{earth} * cos(\phi) - v_{earth}\sqrt{\frac{r_1}{r_1+r_2}} * \sqrt{1-(\frac{r_1}{r_1+r_2} sin(\phi))^2} \\ &= v_{earth} * \Big[cos(\phi) - \sqrt{\frac{1}{\Delta_{au}}(1-(\frac{1}{\Delta_{au}} sin(\phi))^2)}\Big] \end{aligned}$$

□

3.1.4 Proof of lemma 3.1.4

We start with a fixed size and distance, then make limiting arguments over size and distance distributions of get the integral.

One Size Object, One Distance, Many Stars, Many Nights

If there are M 3km objects per square degree at distance R (meters), then the expected number of detections in one light curve is

$$\mathbb{E}\big[\#\text{ of detections}\big] = p * M * f + T\alpha$$

Now suppose we observe L lightcurves over many nights. Each light curve has its own index, even though many are taken during the same zipper run or are of the same star on a different night. For the entire survey over many nights we have:

$$\begin{aligned}
\mathbb{E}\big[\#\text{ of detections in the survey}\big] &= \mathbb{E}\sum_{l=1}^{L}\big[\#\text{ of detections in light curve l}\big] \\
&= \sum_{l=1}^{L}\mathbb{E}\big[\#\text{ of detections in light curve l}\big] \\
&= \sum_{l=1}^{L}\big[p_l * M * f_l + T_l * \alpha\big] \\
&= M * \sum_{l=1}^{L}(p_l * f_l) + \alpha * (\sum_{l=1}^{L} T_l) \\
&= M(3,R) * \Omega_e(3,R) + \alpha * T_{tot}
\end{aligned}$$

Where

$$\Omega_e(D,R) = \sum_{l=1}^{L}(p_l(D,R) * f_l(D,R))$$

And we have

$$\mathbb{E}\big[\#\text{ of detections of (D,R) in the survey}\big] = M(D,R) * \Omega_e(D,R) + \alpha * T_{tot}$$

Many sizes at many distances

Suppose at a fixed distance the population is of objects of size $s_1, ...s_K$. If we were using the same detection statistic for all objects, then we would get addition in the first expectation but not in the one for false positives:

$$\begin{aligned}
\mathbb{E}\big[\#\text{ of detections in survey}\big] &= \mathbb{E}\sum_{i=1}^{K}\big[\#\text{ of real detections of size } s_i\big] + \mathbb{E}\big[\#\text{ of false positives}\big] \\
&= \big[\sum_{i=1}^{K} M(s_i) * eff(s_i)\big] + \alpha * T_{tot}
\end{aligned}$$

A limiting argument would then lead us to

$$\mathbb{E}\big[\#\text{ of detections in survey}\big] = \int M(D)\Omega_e(D)dD + \alpha * T_{tot}$$

Where $M(D)$ has now become the density of objects of size D. The same limiting argument can be done with distance to give

$$\mathbb{E}\big[\#\text{ of detections in survey}\big] = \int M(D,R)\Omega_e(D,R)dD + \alpha * T_{tot}$$

3.2 Inference about population models

Models of the Kuiper belt depend on parameters θ, so $M(D,R) = M(D,R,\theta)$.

Since detections are rare, one can model the outcome of the survey as an observation of a Poisson process with mean

$$\mu(\theta) = \int M(D,R,\theta)\Omega_e(D,R)dD + \alpha * T_{tot}$$

Hence, inference about the mean of a Poisson process can be translated [3] to inference about θ by inverting the function $\mu(\theta)$.

Some confidence limits for the mean of a Poisson variable are [4]

Detections	Lower Limit	Upper Limit
0	NA	2.996
1	.025	5.572
2	.242	7.225
3	.619	8.767

3.2.1 TAOS Results

In [Zhang et al., 2008], constraints are set on power law parameters for KBO's of size $.5km - 10km$ based on zero detections the first two years of data using an algorithm for single point flux reductions. In [Wang et al., 2009] a search for events of longer duration yielded no detections, and constraints are set on power law parameters for several distances. In [Bianco et al., 2010], 3.75 years are examined for KBO's with no detections, and inference is made about many theoretical models which are more complex than a simple power law.

In all three TAOS papers, the false positives $\alpha * T_{tot}$ are not included in calculation of $\mu(\theta)$. The inference remains valid in this case because ignoring the false positive term leads to more conservative results with zero detections. This illustrates the fact that statistical properties of the detection procedure do not need to be known for an upper limit result. However, if detections are present, the false positive rate should be included to get accurate two-sided results.

3.3 Estimation of $\Omega_e(D,R)$

The inference procedures above assume that for all (D,R) we know

$$\Omega_e(D,R) = \sum_{l=1}^{L}(p_l(D,R) * f_l(D,R))$$

[3] Assuming that the function $\mu(\theta)$ is invertible, which requires monotonicity in θ

[4] Some texts report $(0, 3.7)$ for the limit in the case with zero detections; for example, see Table G in [Box et al., 2005]. We consider the the zero detections case a 1-sided test, and the upper limit of 3 is used in all TAOS calculations.

In practice, simulation estimates Ω_e at many points on a grid in (D, R) space. The estimate at each point is unbiased and the standard errors are small. We can then smooth the points to create the function $\Omega_e(D, R)$, or note that the integral $\int M(D, R, \theta)\Omega_e(D, R)dD$ is approximated by a sum over the grid in (D, R) space. We will describe the estimation procedure here and derive bounds on the standard errors.

3.3.1 Estimation for a fixed (D, R)

A simulator has been built [Nihei et al., 2007] to reproduce what an occultation would look like given parameters of the star, the orientation of the telescope, the size of the object, distance to the object, and the impact and timing parameters. One can estimate $\Omega(D, R)$ by either of the following algorithms

Algorithm 3.3.1. *Random Selection*

$for(i \in 1 : N)\{$

- *Pick light curve $l(i)$ uniformly from the L light curves*
- *Generate a time uniformly in that light curve*
- *Generate a impact and timing parameter uniformly*
- *Use the simulator to implant the object in the time series*
- *Apply your procedure to see if you detect the object*
- *Set $X_i = f_{l(i)}\mathbf{1}_{Detect\ during\ i}$*

$\}$

$\hat{\omega} = \frac{L}{N}\sum_{i=1}^{N} X_i$

Algorithm 3.3.2. *Deterministic Selection*

$for(i \in 1 : L)\{$

- *Pick light curve i*
- *Generate a time uniformly in that light curve*
- *Generate a impact and timing parameter uniformly*
- *Use the simulator to implant the object in the time series*

- *Apply your procedure to see if you detect the object*
- *Set* $X_i = f_i \mathbf{1}_{Detect\ during\ i}$

$\}$
$\hat{\omega} = \sum_{i=1}^{L} X_i$

In the random selection case

$$\mathbb{E}(\hat{\omega}) = \mathbb{E}(L * X_1) = L * (1/L \sum_{i=1}^{L} f_l \mathbb{E}(\mathbf{1}_{\text{Detect during i}})) = \sum_{i=1}^{L} f_{l_i} p_l = \Omega$$

In the deterministic selection case

$$\mathbb{E}(\hat{\omega}) = \sum_{i=1}^{L} f_{l_i} p_l = \Omega$$

So our estimate is unbiased in both cases.

To estimate the variance of our estimate, we use

$$\hat{\sigma}_\omega^2 \equiv (\frac{L}{N})^2 * N * \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2$$

Where $L = N$ in the deterministic selection case. It turns out our estimate is either unbiased or conservative:

Lemma 3.3.3. *Variance Estimates*
In the random selection case $\mathbb{E}(\hat{\sigma}_\omega^2) = \text{Var}(\hat{\omega})$
In the deterministic selection case $\mathbb{E}(\hat{\sigma}_\omega^2) \geq \text{Var}(\hat{\omega})$

Proof. Of Lemma 3.3.3 (Random Selection Case)
Since the X_i are i.i.d.

$$\begin{aligned} \text{Var}(\hat{\omega}) &= L^2 * \text{Var}(\bar{X}) \\ &= \frac{L^2}{N} \text{Var}\, X_i \\ &= \frac{L^2}{N} \sigma^2 \end{aligned}$$

And we know that an unbiased estimate for σ^2 is given by

$$\frac{1}{N-1}\sum_{i=1}^{N}(X_i-\bar{X})^2$$

□

For the deterministic selection case, we use

Lemma 3.3.4. *Suppose $X_i \sim (\mu_i, \sigma_i^2) \quad i = 1, ...n \quad$ and suppose the X_i are independent. Then the expected sample variance is the average variance plus the variation in the means:*

$$\mathbb{E}\big(\frac{1}{N-1}\sum_{i=1}^{N}(X_i-\bar{X})^2\big) = \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 + \frac{1}{N-1}\sum_{i=1}^{N}(\mu_i-\bar{\mu})^2$$

Proof.

$$\begin{aligned}
\mathbb{E}(X_i-\bar{X})^2 &= \mathbb{E}\bigg((X_i-\mu_i)+(\mu_i-\bar{\mu})+(\bar{\mu}-\bar{X})\bigg)^2 \\
&= \mathbb{E}(X_i-\mu_i)^2 + \mathbb{E}(\mu_i-\bar{\mu})^2 + \mathbb{E}(\bar{\mu}-\bar{X})^2 \\
&\quad +2\bigg(0+0+\mathbb{E}\big((X_i-\mu_i)(\bar{\mu}-\bar{X})\big)\bigg) \\
&= \sigma_i^2 + (\mu_i-\bar{\mu})^2 + \mathbb{E}(\bar{\mu}-\bar{X})^2 + 2\mathbb{E}\big((X_i-\mu_i)(\bar{\mu}-\bar{X})\big)
\end{aligned}$$

This implies that

$$\begin{aligned}
\mathbb{E}\big(\sum_{i=1}^{N}(X_i-\bar{X})^2\big) &= \sum_{i=1}^{N}\sigma_i^2 + \sum_{i=1}^{N}(\mu_i-\bar{\mu})^2) + N\mathbb{E}(\bar{\mu}-\bar{X})^2 \\
&\quad +2*\mathbb{E}\big((\bar{\mu}-\bar{X})(\sum_{i=1}^{N}(X_i-\mu_i))\big) \\
&= \sum_{i=1}^{N}\sigma_i^2 + \sum_{i=1}^{N}(\mu_i-\bar{\mu})^2) - N\operatorname{Var}\bar{X} \\
&= \frac{N-1}{N}\sum_{i=1}^{N}\sigma_i^2 + \sum_{i=1}^{N}(\mu_i-\bar{\mu})^2
\end{aligned}$$

□

Proof. Of Lemma 3.3.3 (Deterministic Selection Case)

Here we have

$$\operatorname{Var}(\hat{\omega}) = \sum_{i=1}^{L}\operatorname{Var}(X_i) = \sum_{i=1}^{L}\sigma_i^2$$

Combining with the previous lemma gives

$$\mathrm{Var}(\hat{\omega}) \leq N * \mathbb{E}\bigg(\frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2\bigg) = \mathbb{E}(\hat{\sigma}_\omega^2)$$

□

Relationship to weighting coefficients

Suppose we run the loop N times and there are D detections, where we could have $N = L$ in the deterministic setup. We rearrange our indices so the first D are the detected ones. We define $\bar{f}_d = \frac{1}{D}\sum_{i=1}^{D} f_i$. Then,

$$\begin{aligned}\hat{\omega} = \frac{L}{N}\sum_{i=1}^{N} a_i &= \frac{L}{N}\sum_{i=1}^{D} f_i \\ &= L * \frac{D}{N} * \bar{f}_d \\ &= \hat{p} * (L * \bar{f}_d)\end{aligned}$$

Where $\hat{p} = \frac{D}{N}$ is an estimate of our 'Overall Detection Probability' in the case where all light curves have the same length. We also define $\sigma_D^2 = \frac{1}{D}\sum_{i=1}^{D}(f_i - \bar{f}_d)^2$ and let $X_i = f_i \mathbf{1}_{d_i}$.

Lemma 3.3.5. *In this parameterization, our estimate of variance becomes*

$$\begin{aligned}\hat{\sigma}_\omega^2 &\equiv (\frac{L}{N})^2 * \frac{N}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2 \\ &= D(\frac{L}{N})^2\bigg(\frac{N}{N-1}\sigma_D^2 + \frac{N-D}{N-1}(\overline{f_d})^2\bigg)\end{aligned}$$

Where $L = N$ in the deterministic selection case.

Proof. To see that the second equality holds, we rearrange indices so the first D are the detections:

$$\sum_{i=1}^{N}(X_i - \bar{X})^2 = \sum_{i=1}^{D}(f_i - \bar{X})^2 + (N-D)(\bar{X})^2$$

Note that

$$\sum_{i=1}^{D}(f_i - \bar{X})^2 = \sum_{i=1}^{D}(f_i - \overline{f_d} - (\bar{X} - \overline{f_d}))^2 = D\sigma_D^2 + D(\bar{X} - \overline{f_d})^2$$

Then note,

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i = \frac{1}{N}\sum_{i=1}^{D} f_i = \frac{D}{N}\overline{f_d}$$

Putting it together:

$$\begin{aligned}\frac{N}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2 &= \frac{N}{N-1}\left(D\sigma_D^2 + D(\frac{D-N}{N})^2(\overline{c_d})^2 + (N-D)(\frac{D}{N})^2(\overline{c_d})^2\right) \\ &= D\left(\frac{N}{N-1}\sigma_D^2 + \frac{N-D}{N-1}(\overline{f_d})^2\right)\end{aligned}$$

□

We also note that if N is large and we set $p = \frac{D}{N}$ then

$$\hat{\sigma}_\omega = N^{-1/2} * L * \sqrt{p\big(\sigma_D^2 + (1-p)(\overline{f_d})^2\big)}$$

Since our estimate of Ω is $\hat{\omega} = \frac{L}{N} * \sum_{i=1}^{N} a_i = \frac{L}{N} * D * \bar{f}_d = L * p * \bar{f}_d$, we estimate the coefficient of variation as

$$\begin{aligned}\hat{CV} &\approx N^{-1/2} * \left(\frac{p\big(\sigma_D^2 + (1-p)(\overline{f_d})^2\big)}{p^2(\overline{f_d})^2}\right)^{1/2} \\ &= N^{-1/2} * \left(\frac{(\frac{\sigma_D}{\overline{f_d}})^2 + (1-p)}{p}\right)^{1/2}\end{aligned}$$

Hence, we need to do a lot of implantations when the detection probability is small or the coefficient of variation of the (detected) f_l constants is large.

Chapter 4

Filters and Diagnostics

The rank product method assumes that the time series from different telescopes are independent of each other and that each individual series is stationary (and mixing). However, most raw data shows substantial changes in the mean throughout the run and many runs show changing variance. On the one hand, this is a violation because a stationary time series has constant mean and variance. On the other hand, since the changing mean is usually common among the telescopes the time series appears to have low frequency correlation, which violates the independence assumption. At any rate, one must process the light curves to have constant mean and variance before using the rank product.

Figures 4.1 and 4.2 show raw light curves from from a zipper run [1], and another raw light curve from a different zipper run [2] appears in the top panel of Figure 4.4. The light curve sets in Figure 4.1 show that patterns can be common for a star across the three telescopes, and Figure 4.2 shows they need not be common. Figure 4.2 also shows that outliers can occcur, and the top of Figure 4.4 illustrates how the variance can change. Finally we note that the y-axes are not the same for different telescopes, but this lack of calibration does not violate the validity of the rank statistics methodology.

We analyze two univariate light curve filters, and argue that the rank product distribution is not valid for one of them. We also note how autocorrelation is induced by the filter. Next we consider filtering based on other stars and estimate the extent of the possible noise reduction. Finally, we develop hypothesis tests for independence among the filtered, stationary series.

4.1 Univariate Light Curve Filters

We discuss two filters for univariate time series considered in TAOS, the 'Three sigma filter' and 'The EW filter'. We argue that while both force a constant mean, 'The EW filter' induces a changing variance, and therefore should not be used in conjunction with the rank product method.

[1] Field 060, Janrary 13, 2007
[2] Field 060, October 1, 2006

In what follows we start with original data: $x_1, ...x_T$. Also, a window or filter of length $2 * k + 1$ is said to have half-length k.

4.1.1 The Three Sigma Clipping Filter

This filter subtracts off a robust local measure of location, then divides by a robust local measure of spread. It depends on 2 parameters, k_l, k_s, the half lengths of the location and scale filters (For TAOS, $k_l = 16, k_s = 62$). In detail,

1. For $j \in \{k_l + 1, ..., T - (k_l)\}$ let

$$l_j = \text{ Mean of } \{x_{j-k_l}, ..., x_{j+k_l}\}$$

Calculated with a 3 sigma clipping [3] and then set

$$y_j = x_j - l_j$$

2. For $j \in \{k_l + k_s + 1, ..., T - (k_l + k_s)\}$ let

$$s_j = \text{ SD of } \{x_{j-(k_l+k_s)}, ..., x_{j+(k_l+k_s)}\}$$

Calculated with a 3 sigma clipping and then set

$$z_j = y_j / s_j$$

3. (Optional) Take a moving average of resulting z_j

$$z_j^{MA} = 1/3 * (z_{j-1} + z_j + z_{j+1})$$

We note

- For single point flux reductions, the third step is skipped.
- After the first two steps the mean and variance should be approximately constant.
- Other robust measures could be used, such as the median and median absolute deviation (MAD)

[3] A statistic, $T(x)$ is said to be calculated based on the 3 sigma clipping, if it is calculated based on a subset of x that is all within three sigma of the mean. The algorithm used calculates μ and σ for x, then throws out all points with greater than 3 sigma deviation, then re-calculates μ and σ and proceeds until all data is within 3 sigma.

4.1.2 The EW filter

This filter depends on 2 parameters, k_{sig}, k_{bg}, the half lengths of background and signal windows. The parameters considered in [Wang et al., 2009] were $k_{bg} = 10, k_{sig} = 2$.

For $j \in \{k_{bg} + 1, ..., T - (k_{bg})\}$

$$l_j = \text{ Median of } \quad \{x_{j-k_{bg}}, ..., x_{j+k_{bg}}\}$$

$$EW(j) = \sum_{i=j-k_{sig}}^{j+k_{sig}} (\frac{x_i}{l_j} - 1)$$

4.1.3 Comparison

$$EW(j) = \sum_{i=j-k_{sig}}^{j+k_{sig}} (\frac{x_i}{l_j} - 1) = \frac{1}{l_j} \sum_{i=j-k_{sig}}^{j+k_{sig}} (x_i - l_j)$$

Inside the last sum, the l_j is fixed. If we allow it to vary with i, the results shouldn't change too much. Also, the median should be close to the mean with a three sigma clipping, so borrowing some notation from above:

$$EW(j) \approx \frac{1}{l_j} \sum_{i=j-k_{sig}}^{j+k_{sig}} y_i \approx \frac{s_j}{l_j} \sum_{i=j-k_{sig}}^{j+k_{sig}} z_i = \frac{s_j}{l_j} * (2 * k_{sig} + 1) * z_j^{MA}$$

Where the second approximation assumes s_j is roughly constant within the sum.

Since z_j^{MA} has (approximately) zero mean and constant variance, we see that $\mathbb{E}(EW(j)) = 0$ and

$$\sqrt{\text{Var}(EW(j))} \quad \approx \quad \frac{s_j}{l_j} \sqrt{\text{Var}((2 * k_{sig} + 1) * z_j^{MA})} \quad \propto \quad \frac{s_j}{l_j}$$

Unless $\frac{s_j}{l_j}$ is constant, the EW filtered data will have changing variance, and hence will be inappropriate for use with the rank product method.

It should be emphasized that this filter is inadequate even in the Poisson noise case. With Poisson noise $s_j = \sqrt{l_j}$, so the variance is proportional to $\frac{1}{\sqrt{l_j}}$.

Figure 4.3 shows some simulated data designed to emphasize the deficiencies in the EW filter. Figure 4.4 shows a TAOS light curve with both the EW filter and the Three Sigma Filter with optional step three. It can be seen from Figure 4.4 that high and low ranks are more likely later in the EW filtered light curve in , which makes p-values computed from the rank product distribution invalid.

4.1.4 Autocorrelation

The three-sigma filter causes some autocorrelation in the resulting data. This is acceptable for use with the rank product, as was explained in Chapter 2. In Figure 4.5 we see the autocorrelation functions for three filtered light curves, and for a series of filtered Gaussian white noise. The plots show that significant autocorrelation is present, and the shape of the ACF in the real data is similar to what is induced in white noise by the filter.

4.1.5 Adjustments To the Three-Sigma Filter

If a flux reduction occurs and is less than three sigma, then by subtracting the mean the extent of the flux reduction could be suppressed in the filtered data. To avoid this potential decrease in power, we exclude the most central points in calculating location and scale.

It was mentioned in Chapter 2 that the rank product distribution will remain valid if we take a running function of the data. For a general test statistic $T(x_1, x_2, x_3)$, we would run the first two steps of the three sigma filter, excluding the three central points in the calculation and then use T in the optional step 3. If T were a function of 20 points, we would need to rethink our filtering methods, since the mean window is 33 points. As we will show in Chapter 9, the signals of interest for TAOS are all less than 5 points long at opposition. [4]

4.2 Filters Based On Other Stars

Searching for occultations in TAOS data means looking for signals of length $1-5$ points in series of length 27000, and the Three-Sigma Clipping filter is capable of removing the background signal without erasing the occultation signal. If we were interested in longer signals, we could exploit the fact that some signals are common on many stars [5], and use regression to remove those components. In [Kovacs et al., 2005] regression is used to decrease noise while searching for periodic transients. In [Kim et al., 2009], a algorithm using clustering and regression is used to improve detection efficiency of simulated signals whose length is of the order of the time series.

Although the occultation search in TAOS does not need to filter using other stars, it is natural to ask whether detection efficiency can be improved using regression on other stars. Some time was spent on this problem, particularly in searching for objects in the Kuiper Belt region, but we had little success. We found that such filtering resulted in small increases in the power to detect implanted occultations, making the increased computational cost and potential spread of outliers hard to justify.

Here we give a heuristic argument suggesting [6] that regression might decrease the noise level by 10%, and in Chapter 6 we show how such a noise reduction would cause a very small increase power in the case where errors are Normal. Here we are only concerned with

[4] Most TAOS observations are at opposition. With a bigger elongation angle, the signals become longer.
[5] The main reason is that they are passing through the same atmosphere.
[6] A sophisticated modeler might find a way to get a greater noise reduction out of this data.

approximating potential noise reduction, and not the negative side effects of the procedures considered.

4.2.1 Heuristic Regression Argument

On a given telescope, global changes in the mean and variance may be common among some stars, but can look very different on other groups of stars. Since we are not interested in global trends, we use the Three-Sigma filter on all stars giving the plots in Figure 4.6, and use regression to remove the remaining correlation. For the positive outlier around time 7000 on Star 16 of Figure 4.6, there are no other outliers at that time point on any of the telescopes. For the negative outlier around time 26000, the outlier appears on over 30 stars on telescope D, but never on the other telescopes. We gloss over issues relating to dealing with outliers [7] by considering two data sets: one is the filtered data as it appears in Figure 4.6 (called RAW in plots), and one where all points greater than 6 median absolute deviations are set to zero.

Suppose Y is an $N \times p_1$ matrix of p_1 light curves and suppose X is an $N \times r$ matrix of signals and we want to filter based on the multivariate linear model

$$Y = X\beta + \epsilon$$

Where column j of ϵ has variance σ_j^2.

We ignore the fact that autocorrelation is present in the real data, and use $\hat{\sigma_j^2} = \frac{1}{N-r}||e_j||^2$ as an unbiased estimate for σ_j^2, where e_j is the N-vector of residuals for the j^{th} light curve after fitting the linear model.

We do not want to regress a light curve on itself, and there are computational limitations, so one possibility is to split the data into two groups, let Y be one of the groups, then let X be the first r principal components of the other group of light curves.

For the zipper run of Figure 4.6, we split the data and in the top of Figure 4.7 we look at the average (over the 1/5) estimate of variance as a function the number of principal components used. This result suggests that for error reduction, we should regress on all 4/5 of the other light curves.

Next we do 20 iterations of

1. Randomly select 20% of the light curves.
2. Regress the 20% on the other 80%.
3. Calculate the Residuals, then the estimated error variance.
4. Save the square root, because the standard deviation is the important quantity.

In the bottom of figure 4.7 we look at density plots for the estimate of the standard deviation of the regression filtered data. We note that

[7] The outliers of greatest concern are negative outliers that occur simultaneously on all three telescopes. This happens frequently, but when it does it occurs on many stars at the same time point, indicating that it is a photometry artifact. As part of the detection pipeline, such time points are deleted from the data set. More details about the detection pipeline appear in [Bianco et al., 2010]

- The error reduction on telescope A for the raw data is almost the same as when outliers are clipped.
- Telescope D appears to respond a little better to the procedure.
- Although a handful of light-curves respond well, the vast majority decreased less than 10%

The same experiment was performed on other zipper runs, and in all cases claiming an approximate 10% noise reduction seems generous.

4.3 Diagnostics

We consider the *Chi-Square Test* to test for independence between multiple series. The approximate distribution of the test statistics is known analytically for iid data, but breaks down for stationary data. We saw in Figure 4.5 that filtered data has autocorrelation, so the methods need to be generalized.

We introduce the *Circular Shift* and the *Block Bootstrap* as tests for independence between multiple stationary time series.

4.3.1 Chi Square Test

Suppose we have K time series of length N and $N = G * M$. We arrange the rank tuples into a 'cube' and divide each coordinate into G groups of length M, so there are G^K groups, each containing M^K points. Let O_i be the number of rank tuples in cube i. Then

$$E_i \equiv \mathbb{E}(O_i) = \frac{N}{G^K}$$

and the χ^2 statistic is given by

$$T = \sum_{i=1}^{G^K} \frac{(O_i - E_i)^2}{E_i}$$

The degrees of freedom is G^K minus the number of independent constraints.

First, we note that the total number of points is N, giving one constraint. In the case of $K = 1$, the degrees of freedom is $df(1) = G-1$. For $K = 2$ we get a grid in rank-rank space, and each column must have exactly M points, giving another G constraints so that $df(2) = G^2 - G - 1$. For each additional dimension we add, we get another G constraints, so our general formula is

$$df(K) = G^K - G(K-1) - 1$$

If the original data is iid, then we have the $T \sim \chi^2_{df(K)}$. However, if the original data is stationary then this will not necessarily hold.

Suppose we have positive autocorrelation in the series. The E_i remains the same in the expression for T, but there is likely to be more clumping so that the variability of O_i around its mean will be bigger, implying that the statistic T will be bigger than in the iid case.

4.3.2 Stationary Versions of the Chi-Square Test

For ease of discussion and notation, we assume we have three series of length 26000, and the grid for the chi-square statistic is size 5; $K = 3 \;\; N = 26000 \;\; G = 5$.

The Block Bootstrap

Here we do a stationary analog of a permutation test called the Block Bootstrap [Kunsch, 1989]. The idea is that dependance between series will be broken by the permutation of blocks, but the marginal (stationary) distributions will remain basically the same.

Along each of the three axis, divide into 10 blocks of length 2600. Next, generate three random permutations of $1, 2, ..., 10$, independently, and reorder the blocks along each axis, then calculate the Chi-Square statistic.

For added clarity: if you generated a permutation $(2, 1, 6...)$ then you would get

$$2601, 2602, ..., 5200, 1, ...2600, ..., 13001, ...15600, ...$$

We calculate the Chi-Square statistic on the original data, then generate 99 new statistics by permuting the blocks. We then look at the rank of the chi-square on the original data in the list of 100 values r_1 and get a p-value $p = \frac{r_1}{100}$

The Circular Shift

Here we calculate the chi-square statistic and compare it to shifted versions of itself, with shift increments $(0, 260, 780)$.

Letting f be the chi-square statistic, this gives:

$$g_0 = f\Big((x_1, ..., x_{26000}), (y_1, ..., y_{26000}), (z_1, ..., z_{26000})\Big)$$

$$g_1 = f\Big((x_1, ..., x_{26000}), (y_{261}, ..., y_{26000}, y_1, ..., y_{260}), (z_{781}, ..., z_{26000}, z_1, ...z_{780})\Big)$$

$$g_2 = f\Big((x_1, ..., x_{26000}), (y_{521}, ..., y_{26000}, y_1, ..., y_{520}), (z_{1561}, ..., z_{26000}, z_1, ...z_{1560})\Big)$$

Continuing this way using modular arithmetic will give $g_{100} = g_0$ and there will be no repeats in that cycle of 100. If the original series are SSC and independent of each other, then the series $g_0, ..., g_{99}$ will be SSC. Therefore, we can let r be the rank of g_0 in those 100 numbers and get a p-value $p = \frac{r_1}{100}$.

Comparisons

We consider three marginal distributions; iid, Gaussian $MA(5)$, and Gaussian $MA(15)$, and apply three tests; Upper tail of Chi-Square with $df(3)$ degrees of freedom, the Circular Shift, and the Block Bootstrap. Since we generate independently on the three telescopes, a valid test for independence would give uniform p-values, so a plot of the empirical cumulative distribution function (*ecdf*) would look like the identity function.

Figure 4.8 shows the *ecdf* for the nine cases considered. You can see that the Circular Shift and Block Bootstrap work in all cases considered, whereas using the Chi-Square distribution fails for the moving averages. In the iid case, the correlations between the p-values for the methods is always greater than 0.97. For the moving average cases, the correlation between the p-values for the Circular Shift the Block Bootstrap are also greater than 0.97.

Because of its simplicity, we prefer the Block Bootstrap method and use it in [Lehner et al., 2010].

An Example

In figure 4.9 we see filtered lightcurves [8] for a light curve triple showing dependence. The block bootstrap was used with 1000 iterations and the p-value was $\frac{1}{1000}$.

[8] Star 67, Field 060, on January 13, 2007

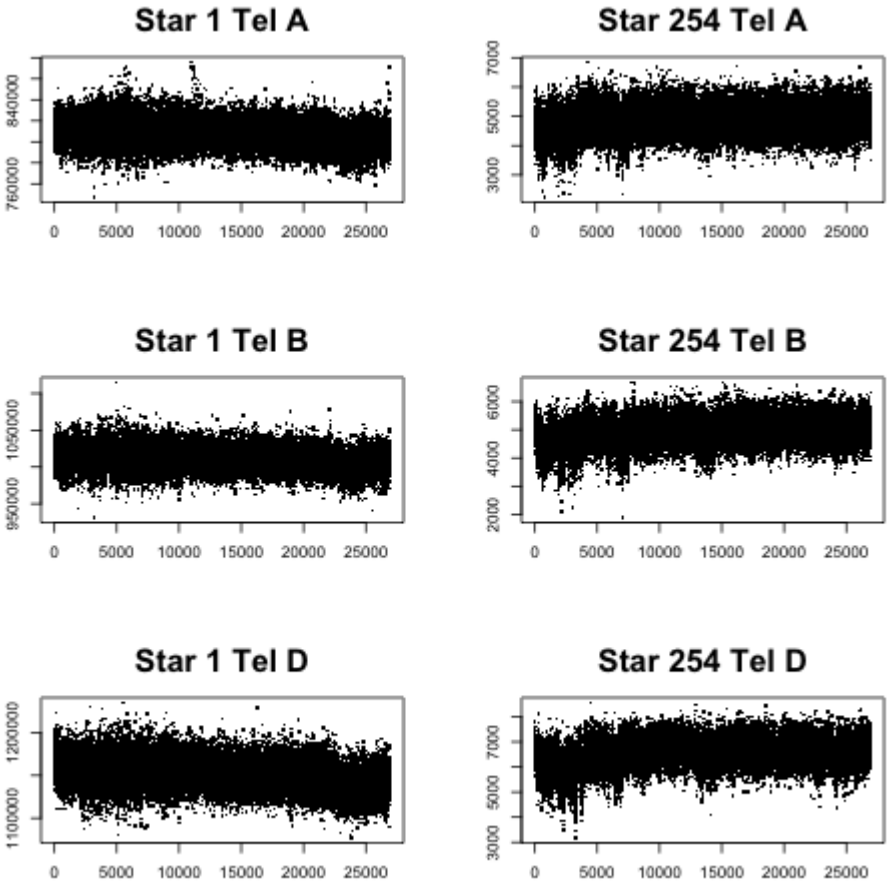

Figure 4.1: Raw Light Curves for two stars on the three telescopes

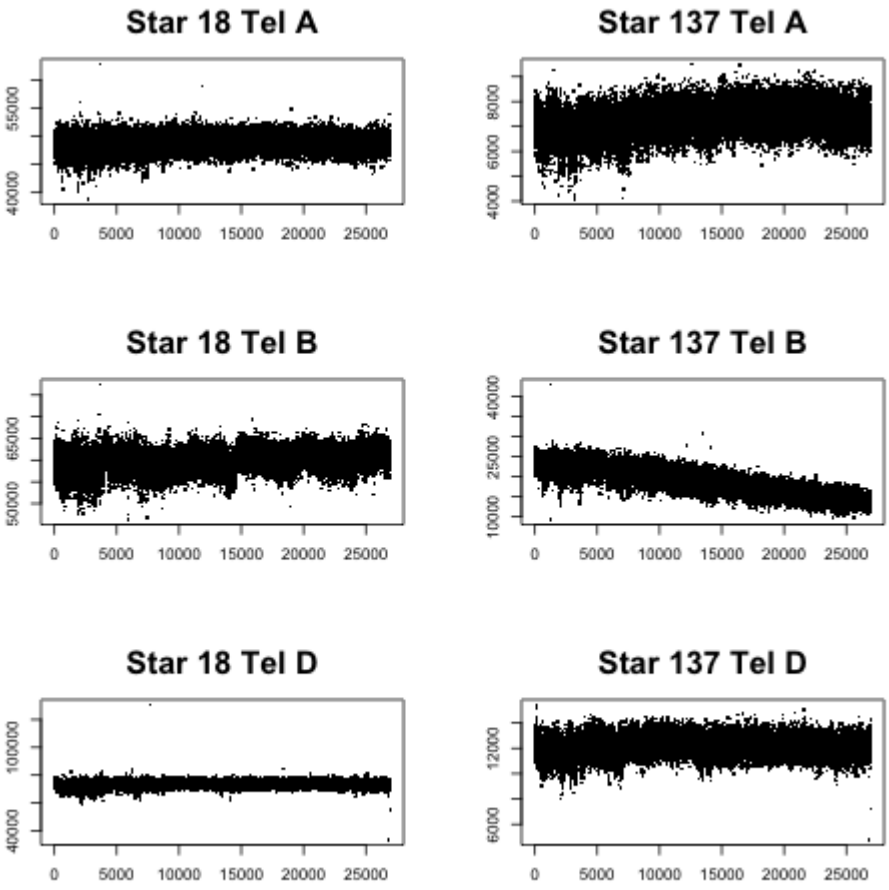

Figure 4.2: Raw Light Curves for two stars on the three telescopes

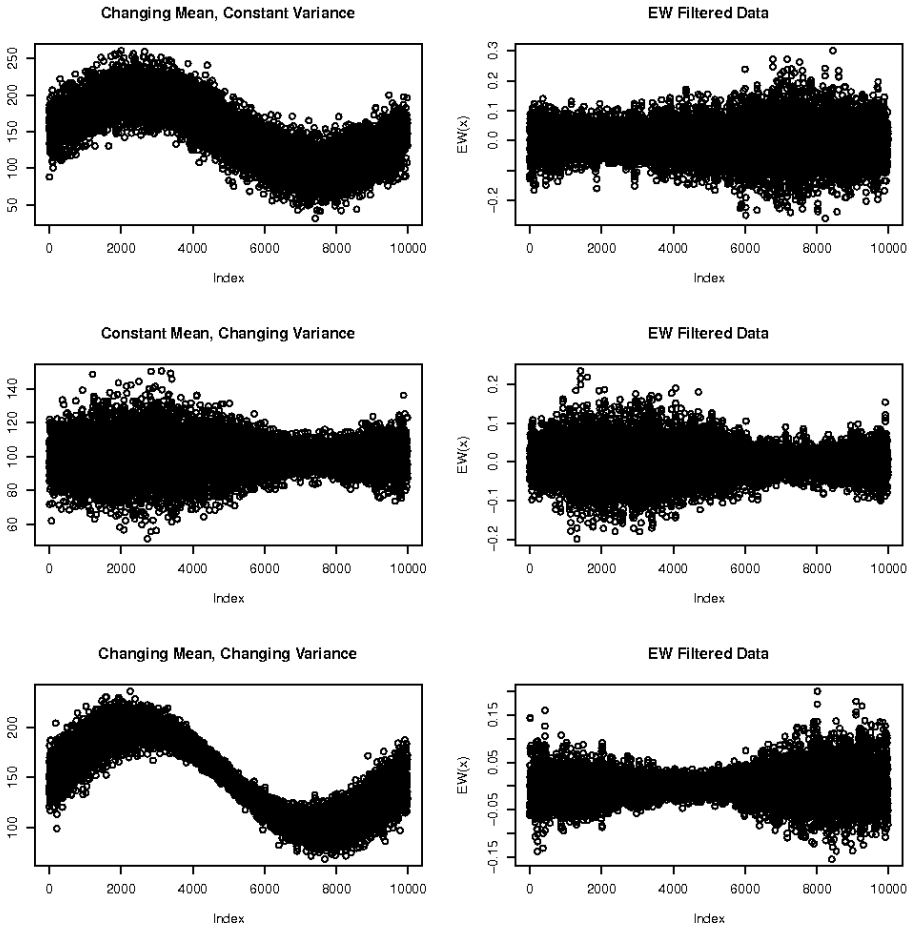

Figure 4.3: Simulated data appears in the left hand column. The right hand column is the result of running the EW filter on the data in the left hand column.

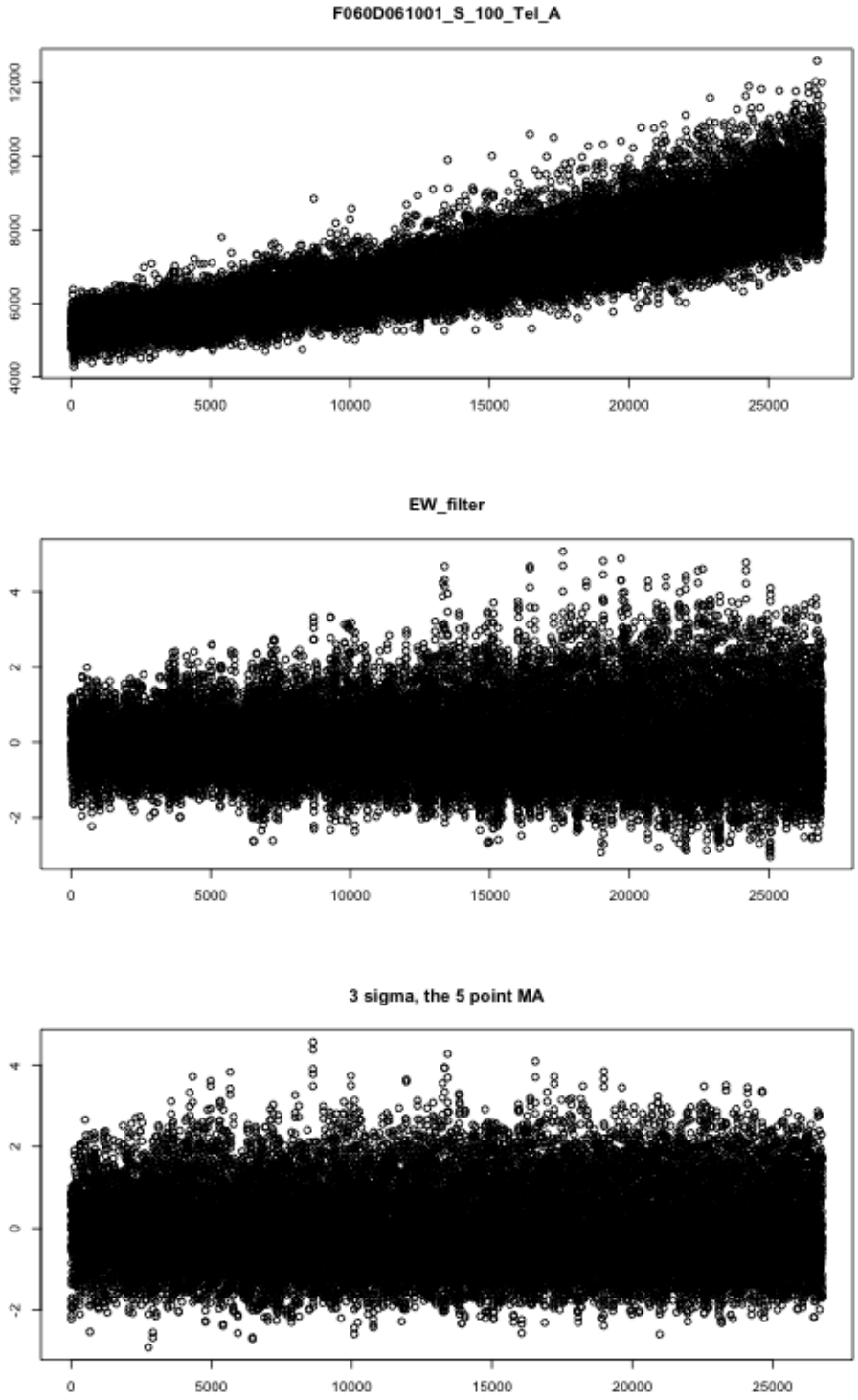

Figure 4.4: Top is raw light curve for Star 100 on Telescope A of Field 060 on October 1, 2006. The other plots are the result of running the EW filter, and the Three-Sigma filter with a 5 point moving average.

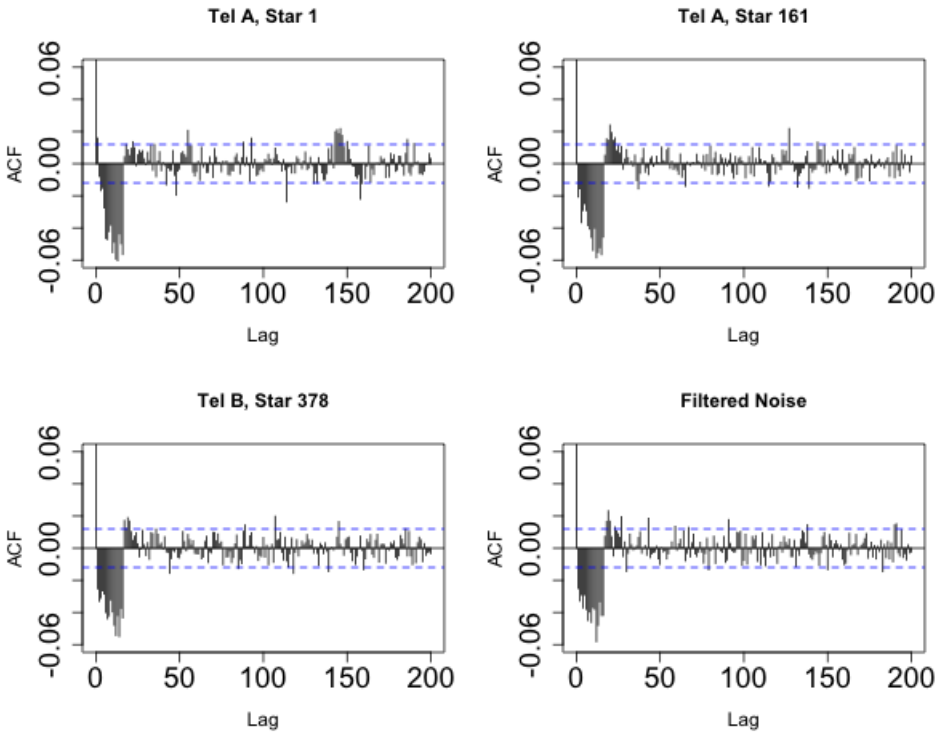

Figure 4.5: Autocorrelations for three filtered lightcurves. The bottom right is the result of running the filter on a white noise series.

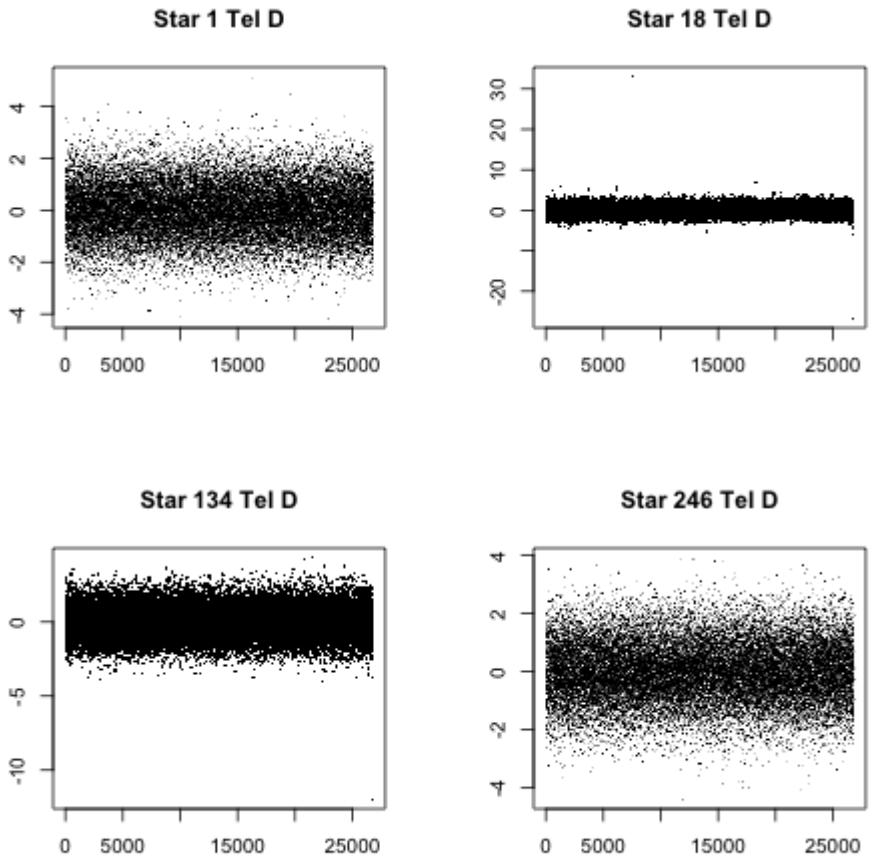

Figure 4.6: From the four stars considered in Figures 4.1 and 4.2, we look at the filtered versions of the light curves for telescope D. The filtered data was slightly adjusted to have median 0 and mad 1.

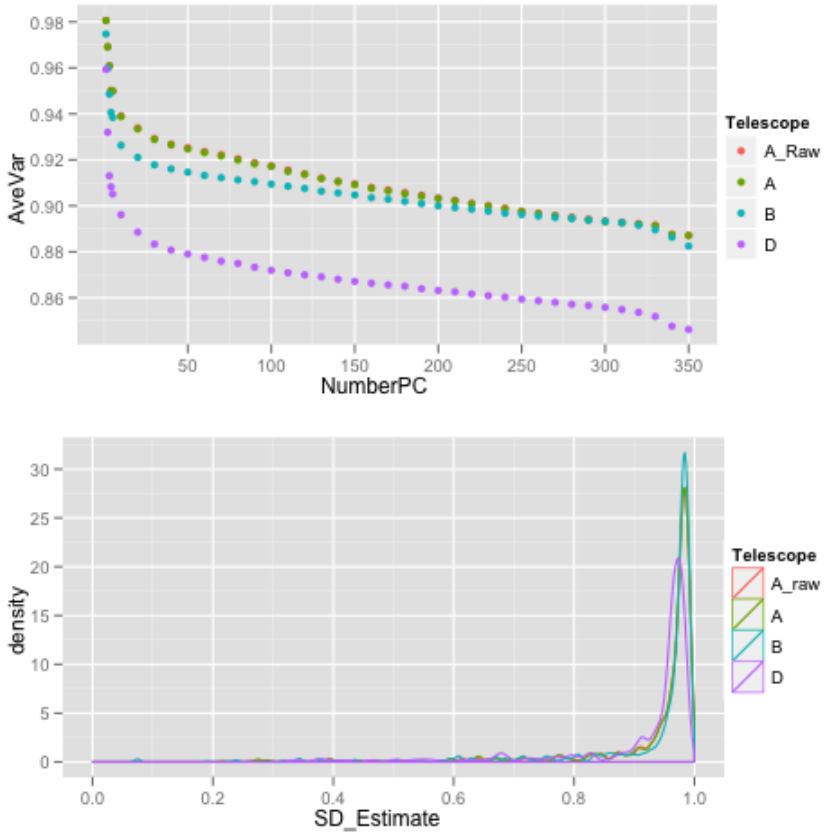

Figure 4.7: **Top Panel:** For a given $(1/5, 4/5)$ partition we look at the average estimate of residual variance (over the $1/5$) as a function of the number of principal components used. Note that A_{raw} is the data on telescope A where we have not clipped the outliers. **Bottom Panel:** We iterate the procedure above several times and keep the estimated SD for each star considered, based on all the principal components.

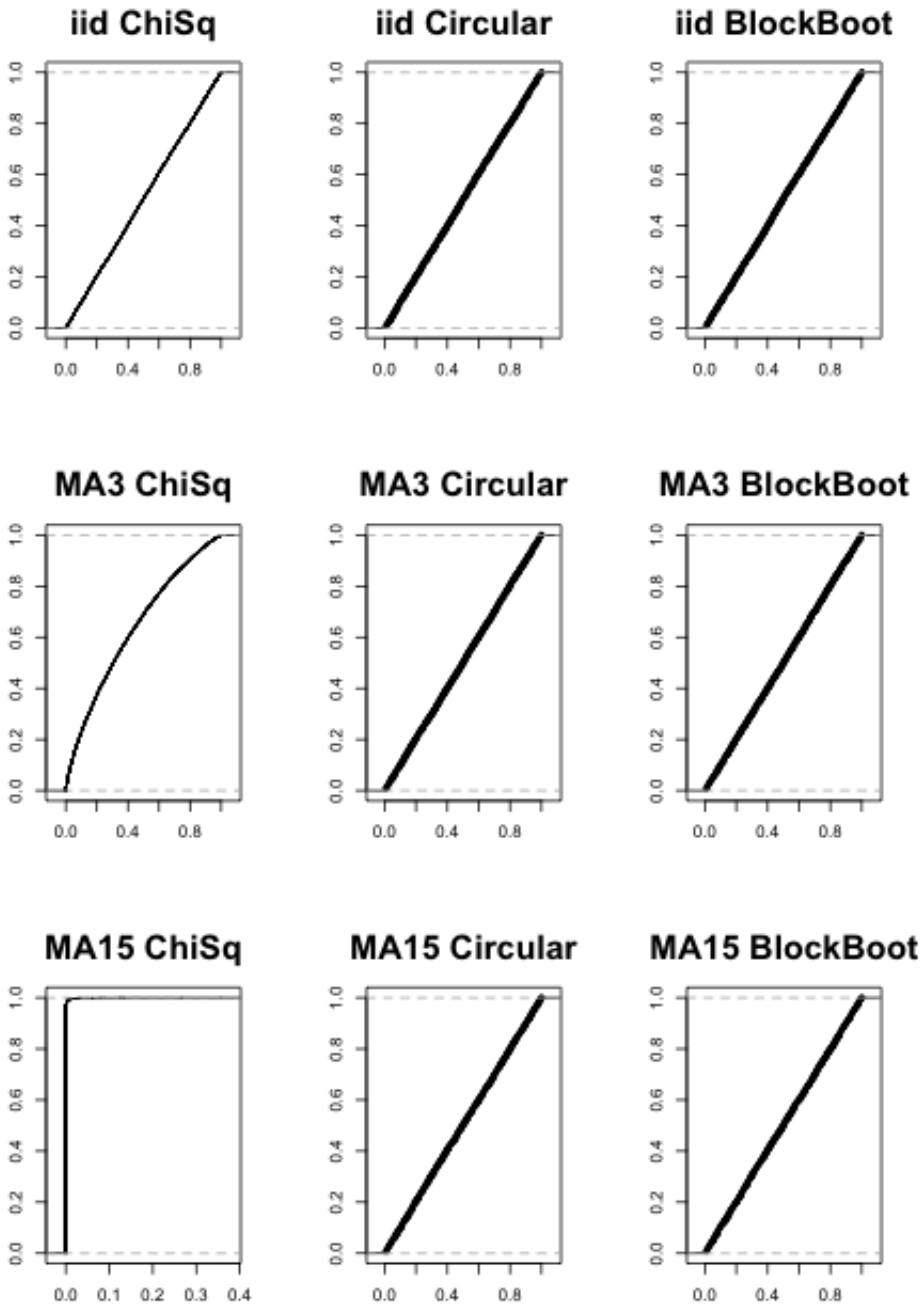

Figure 4.8: Empirical Distribution Functions of p-values for the three test statistics under the three alternative distributions.

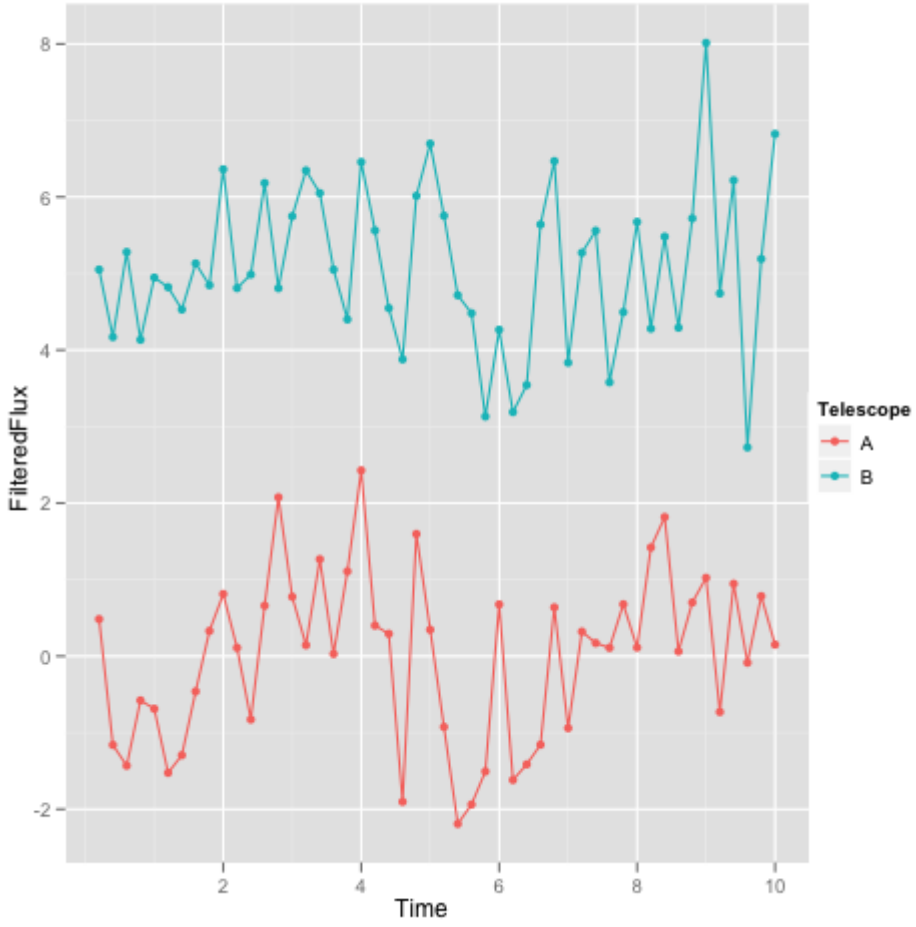

Figure 4.9: Telescopes A and B for a 10 second run in a light curve triple showing dependance by the block bootstrap test. To improve visualization, 5 was added to the filtered flux for telescope B.

Chapter 5

Combination

The rank product method corresponds to a particular rejection region in rank space. As an alternative to the rank product, in [Liang et al., 2004] the maximum rank was considered. Let $r_{i,j}$ be the rank at time i on telescope j in time series of length $N = 27000$ where there are $K = 3$ telescopes. Then the maximum rank s_i is given by $s_i = \max_j r_{i,j}$ and levels can be determined by

$$\begin{aligned} \mathbb{P}(s_i \leq c) &= (\mathbb{P}(r_{i,j} \leq c)^K \\ &= (\frac{\lfloor c \rfloor}{N})^K \end{aligned}$$

This jumps past level 10^{-10} as c goes from 12 to 13, so the threshold is 12. We also recall that at level 10^{-10} the rank product threshold is 125. This implies that the rank tuple $(6, 6, 6)$ would be detected by the maximum rank statistic, but not by the rank product, since $6^3 = 216 > 125$. On the other hand $(1, 1, 13)$ would be detected by the rank product but not by the maximum rank. Which method is better?

In this chapter we compare the maximum rank with the rank product on simulated data. Based on the simulation, we conclude that the rank product is better than the maximum rank, and that there isn't much room for improvement.

5.1 Simulated Occultations with TAOS data

We implanted simulated occultations of a 3km object at 43 AU in a zipper [1] run, and kept the corresponding rank-tuples.

More specifically we iterate the following process $4 * 10^5$ times, and we will refer to each iteration as an *implantation*

1. Pick 1 of the 548 stars with probability proportional to the *weighting factor* [2].

[1] The zipper run was Field 060 on October, 1, 2006. The opposition angle was -49. There were 548 light curve sets, and they were of length 26788 after filtering.

[2] We use this sampling scheme so the probability of recover corresponds to the probability of detecting an object randomly distributed in space. We note that for stars on the same zipper run, of the same spectral

2. Get the corresponding H which sets the range of impact parameters.

3. Uniformly generate timing and impact parameter, and create the diffraction profile.

4. Uniformly generate a time in the light curve, and implant the signal there.

5. Run the filters on all the light curves [3].

6. Rank the data on each light curve then save the rank triple corresponding to the time point where the implantation was done [4].

7. Also save the midpoint of the diffraction profile.

In figure 5.1 we plot results from 10^5 iterations and stratify by $a = \%$ flux reduction at midpoint of diffraction profile. The value $a = .2$ means a 20% flux reduction at the midpoint of the signal, and a histogram of the a values appears at the top of the figure. In figure 5.2 we look at the three marginal distributions of ranks for the three telescopes.

From this experiment, we note

- Occultations can cause *increases* in flux. The reason for this is that a kind of constructive interference happens for large impact parameters. As seen in the histogram at the top of figure 5.1, 31.6% of the implantations resulted in flux increases for this experiment. However, the magnitudes of the flux increases are not as big as the flux decreases.

- The increase in flux (negative a), shows up in rank space. In figure 5.1, there is a concentration of points in the upper corner, and it turns out that the highest rank tuple, $(26787, 26787, 26787)$ is the 15^{th} most popular tuple under the alternative. In figure 5.2 you can see that the higher ranks have a higher density, although this mass is small compared to the low ranks.

- From the 3-D plots, it appears that the region from Fisher's method is more reasonable than the box constraint. At level 10^{-10}, one can calculate that the rank product has power .343 while the maximum rank only had power .294. The highly significant difference of 4.9% can be broken down further: under the alternative, the rank product detects when the max does not about 5.0%, and the max detects when the rank product does not about 0.1% of the time.

---

type, the weighting factor is proportional to H, so this sample includes more bright stars than a simple random sample of the stars.

[3] We can speed up this computation by saving a filtered light curve. For each implantation we filter only in the neighborhood of the implantation, then paste that part to saved filtered light curve before ranking.

[4] Another possibility would be to take the minimum of the rank products in a neighborhood of the implantation, since we are considering procedures with low rank. We have found that when there is a flux increase at the midpoint of the implantation, then taking the minimum will give something with slightly lower rank, but not so low to be significant according to the rank product or maximum rank. In the method used, we see how the central part of the signal shows up in the rank-rank-rank space.

- Although the shapes of the marginal distributions are similar, they are not exactly the same. From figure 5.2, it appears that telescope B has less small ranks. Focusing on rank 1, the probabilities are $(0.326, 0.261, 0.343)$, and these differences are all highly significant.

We have considered the maximum rank and the rank product, but any set of 1922 rank-triples could serve as the rejection region for our test because ranks are uniform under the null and $\frac{1922}{26787^3} \approx 10^{-10}$. The Neyman-Pearson lemma implies that the optimal region is the 1922 rank triples with the highest probability under the alternative.

The *Optimal Power* can be estimated by simulations, but statistical properties are hard to access because unlike the rank product or maximum rank, the region is data dependent. However, we can make statistical claims about the *difference* between the rank product and the optimal region.

Lemma 5.1.1. *Confidence Regions*
Suppose R_1 and R_2 are regions with probabilities p_1 and p_2 under the alternative. Suppose we draw N iid copies under the alternative, and count the proportions falling in the regions, giving $\hat{p_1}$ and $\hat{p_2}$. Then for large N we have,

$$\mathbb{P}\left((p_2 - p_1) \leq (\hat{p_2} - \hat{p_1}) + \frac{1}{\sqrt{N}} \Phi^{-1}(1-\alpha) \right) \geq (1-\alpha)$$

Proof. Let $z_i = \mathbf{1}_{T_i \in R_2} - \mathbf{1}_{T_i \in R_1}$. Then $\bar{z} = (\hat{p_2} - \hat{p_1})$ and $\mathbb{E}(\bar{z}) = \mathbb{E}(z_i) = (p_2 - p_1)$. We assume we know $\sigma_d^2 = \operatorname{Var} z_i$. Because N is large, we can make the normal approximation

$$\bar{z} \stackrel{d}{=} (p_2 - p_1) + \frac{\sigma_d}{\sqrt{N}} \epsilon \qquad \text{Where} \qquad \epsilon \sim \mathcal{N}(0,1)$$

Which gives us

$$\mathbb{P}\left((p_2 - p_1) \leq (\hat{p_2} - \hat{p_1}) + \frac{\sigma_d}{\sqrt{N}} \Phi^{-1}(1-\alpha) \right) \geq (1-\alpha)$$

Since z_i has support on $\{-1, 0, 1\}$, the highest variance it can achieve puts half of its mass on each ± 1, which gives variance of 1. This implies $\sigma_d \leq 1$, which gives the result. □

In an experiment with $4 * 10^5$ implantations, if we select the 1922 triples with the highest count, we get a detection probability of $p_2 = 0.35$, which is larger than the rank product by 0.0066. This implies that for any of the $\binom{26787^3}{1922}$ regions, we would get the conservative 99.9% confidence bound:

$$p_2 - p_1 \leq 0.01147$$

So we conclude that

- Even though there are flux increases that the rank product cannot detect, and even though there is statistically significant asymmetry under the alternative, it appears there is not much to be gained by considering more complicated regions for 3km objects at 43 AU [5].

[5] We also did an experiment with 1km objects and got a confidence bound for the difference less than

1%. In that case the *relative* potential gain was bigger, because the probability of detection by the rank product was about 4.4%. However, the gain was still small, and most of it came from the flux increases, and practically we would not report a flux increase to the scientific community as an occultation.

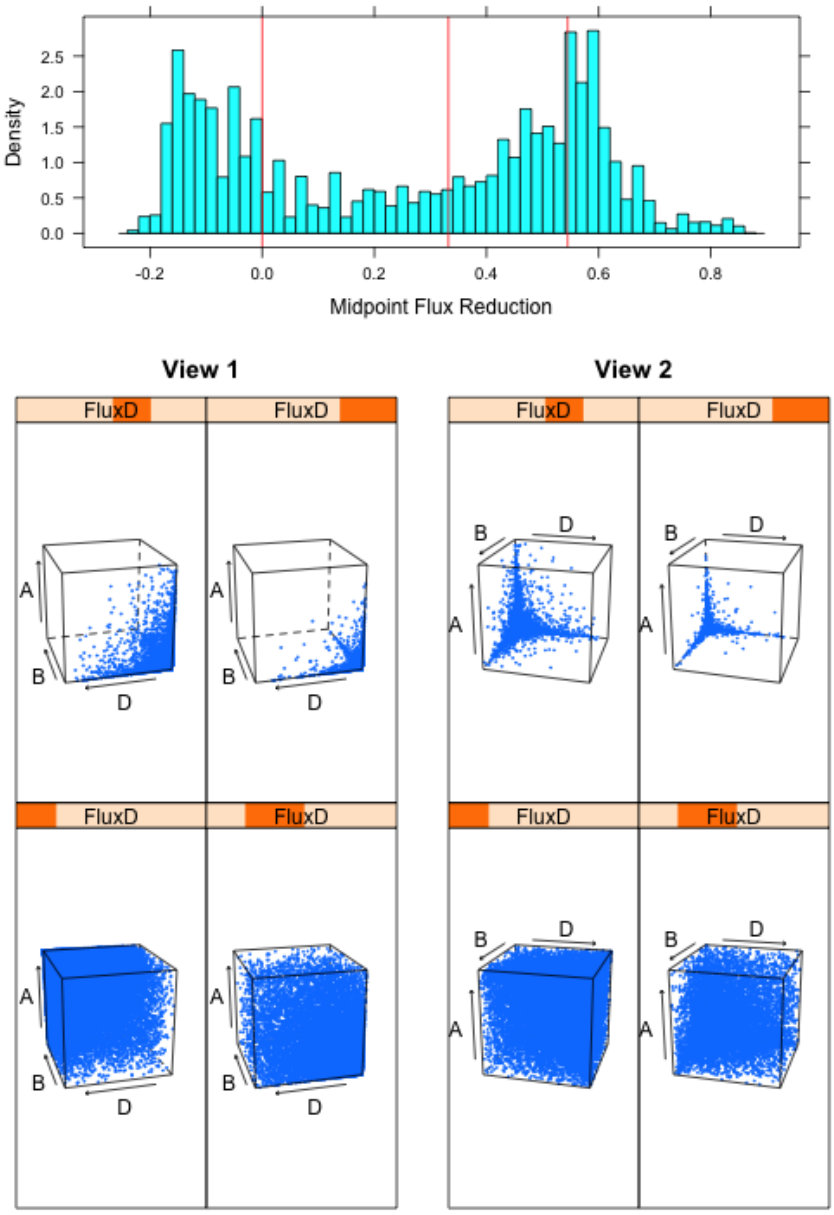

Figure 5.1: 10^5 implantations made in zipper run F060D061001. The histogram on top shows the midpoint flux reductions, and the red lines indicate the groupings of the variables that appear in the bottom plots. For 31.7% of the data we have $a < 0$, which means a flux increase. The other lines are for the median and the 75^{th} percentile.

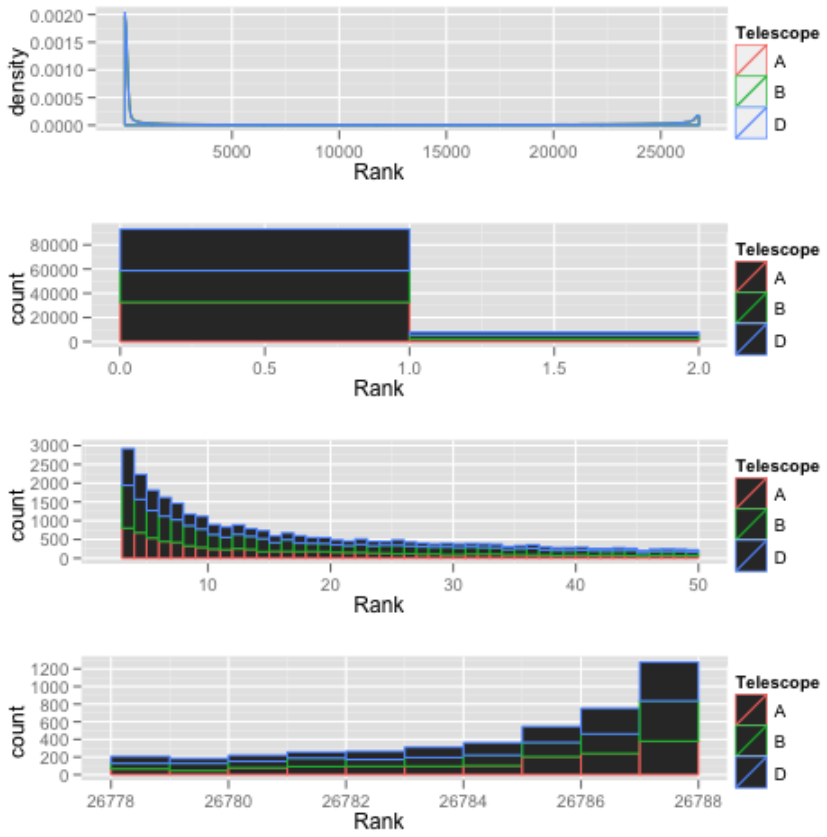

Figure 5.2: **Top:** Density plots for the ranks under the alternative for 10^5 simulations. **Lower Three Plots:** Stacked histograms for the three telescopes for ranges of ranks $1-2$, $2-50$, $26779-26788$.

Chapter 6

Single Point Detection

In this chapter we explore the parametric model of single point flux reduction:

$$y = (1 - a)\mu + \sigma\epsilon$$

Here a is the percentage flux reduction, μ is the intensity of the star, and ϵ is a noise term with mean zero and variance 1. In figure 6.1 we consider the Normal distribution and centered and scaled versions of double exponential distribution, t distribution with 3 degrees of freedom, and negative of a chi square distribution with 5 degrees of freedom. Clearly the non-normal distributions have heavier tails, with t_3 being the heaviest.

We find it convenient use $\lambda = \frac{\mu}{\sigma}$ and transform according to $z = \frac{y-\mu}{\mu}$, which gives

$$z \stackrel{d}{=} -a + \frac{1}{\lambda}\epsilon$$

Letting F be the CDF of the error term, we see that a level α test is given by $z \leq \frac{1}{\lambda}F^{-1}(\alpha)$, and the power of that test is given in terms of the parameter $c = a * \lambda$

$$\begin{aligned} POW(c, \alpha) &= POW(a, \lambda, \alpha) \\ &= F\Big(F^{-1}(\alpha) + a\lambda\Big) \\ &= F\Big(F^{-1}(\alpha) + c\Big) \end{aligned}$$

If the errors are Normal, then given K independent versions of this test, the Neyman-Pearson lemma tells us to co-add the observations, which is equivalent to a single observation with the signal to noise ratio improved by a factor of $\sqrt{K}$. Therefore, the optimal power for independent multi-telescope data in the gaussian case is given by

$$POW(c, \alpha) = \Phi\Big(\Phi^{-1}(\alpha) + \sqrt{K}c\Big)$$

On the other hand, the rank product method depends on the distribution of the rank of the first element of

$$\left(-a + \frac{1}{\lambda}\epsilon_1 \, , \, \frac{1}{\lambda}\epsilon_2 \, , \quad \ldots \, , \quad \frac{1}{\lambda}\epsilon_N \right)$$

Where ϵ_i is the noise sequence. We note that this is equivalent to the distribution of the rank of the first element of

$$\left(-(a * \lambda), \, \epsilon_2 - \epsilon_1 \, , \quad \ldots \, , \quad \epsilon_N - \epsilon_1 \right)$$

So the distribution of the rank product under the alternative also depends on the parameter $c = a * \lambda$.

This leads us to ask:

1. If the data is Normal, how does the power of the rank product compare to the optimal power implied by the Neyman-Pearson lemma for various values of the parameter c?
2. How does the rank product test perform in the 3 and 4 telescope cases, for various values of the parameter c?
3. How does the power change when using the wider tailed distributions in figure 6.1, for various values of the parameter c?
4. For the experiments in chapter 5, what was the power as a function of the implied c parameter, and how does that compare to the other error distributions?
5. For TAOS data, what are typical signal to noise ratios, and how do the distributions of a's relate to size, distance, and other survey parameters?
6. For TAOS data, using the Gaussian model, how is power related to size, distance and other parameters of the survey, and are there design implications?
7. Since we saw in chapter 4 that regression based filtering can lead to a 10% reduction in noise, we ask how a 10% noise reduction changes the power of TAOS assuming the Gaussian model.

6.1 Power in terms of c

The power using the Neyman-Pearson lemma can be calculated theoretically, and the power based on the rank product can be estimated by simulation. By simulating 10^5 times, the standard errors on our probability estimates must be less than $(1/4 * 10^{-5})^{.5} = .0016$. Our simulations use lengths of $N = 27000$, and level $\alpha = 10^{-10}$. In figure 6.2 we look at the power curves for several error distributions and for 3 and 4 telescopes. We note that:

- The Normal test based on Neyman-Pearson is more powerful in the 4 telescope case than the three telescope case; this must be true from theory, and we can see it in figure 6.2.

- From figure 6.2, we see that the rank product test is more powerful for 4 telescopes, for all noise distributions considered.
- In terms of power, we have the hierarchy

$$\text{Normal} \quad > \quad \text{Double Exponential} \quad > \quad \text{Chi Square} \quad > \quad t_3$$

  Which mirrors the ordering of tail behavior in Figure 6.1.
- For all noise distributions, the power is essentially zero for $c < 2$.
- For the Normal Neyman-Pearson and the Normal ranks, there is very high power for $c > 6$.
- The t_3 distribution does quite poorly, especially in the three telescope case.

In figure 6.3 we stick to the three telescope case and also look at real data from two experiments like in Chapter 5. **RealData1** is the experiment explained in Chapter 5; we assume a $3km$ object is uniformly distributed the space that a certain zipper run covers, and we calculate the probability that our procedure recovers the event by doing $4 * 10^5$ implantations. We also save the $4 * 10^5$ corresponding values of $c = \lambda * a$, where λ is the average signal to noise ratio on the three light curves and a is the % flux reduction at the midpoint of the implanted signal. For **RealData2**, we use a different zipper run [1] and implant a $1km$ object. Figure 6.4 shows the distribution of c parameters for the two Real Data experiments.

The standard errors for the real data curves are bigger than the others and are not uniform because some c's are not as likely to occur, as can be seen in Figure 6.4. Still, the usual estimate of standard errors always [2] gives values less than 1.1% for figure 6.3. We note that

- The two real data experiments give very similar power functions.
- The real data is not as powerful as the normal distribution, but it is more powerful than all the other simulated distributions.
- The major differences between the optimal gaussian, the gaussian ranks, and the real data curves appear in the region $c \in (2.5, 5.5)$.
- Much of the distribution of c for the two experiments lies outside the region $(2.5, 5.5)$, as can be seen in Figure 6.4.

[1] This zipper run is of Field 060 on November 2, 2006.

[2] When the estimate of p is 0 or 1, this standard estimate of the SE is 0, which is not correct. Both real data experiments had zero recoveries for $c \leq 1.5$, and this was based on a ton of data as you can see in Figure 6.4. Both experiments also recovered 100% of events where $c \geq 10$, and recovered over 99% for $c \geq 7.25$. For the $1km$ experiment, $\frac{1301}{1301}$ of implantations with $c \geq 7.25$ were recovered, and over 500 of those had $c \in (7.25, 8.25)$.

6.2 TAOS data and implications about power

We have looked at power as a function of $c = \lambda * a$, and found that events with $c < 2$ are not detectable. Since the percent flux reduction, a, must be less than one, there will be **no power** to detect occultations at level 10^{-10} if $\lambda < 2$. More generally, given the signal to noise ratios in the TAOS survey, and using the simulator to produce the values of a, we can understand better where TAOS is likely to be powerful.

In figure 6.5 we see the distribution of signal to noise ratios and magnitudes in the survey, while figure 6.6 shows the signal to noise ratios for three groups of magnitudes: < 10 , $10 - 12$, $12 - 14$.

We note that

- (83%) have magnitudes 12-14, while about 15% have magnitudes 10-12, and about 2% are brighter than magnitude 10. That is, most stars are dim, some are in the medium range, and a negligible number are bright.
- (18%) of stars have signal to noise less than 2, meaning that portion of the data is not likely to recover anything. Almost all of these occur for stars in the dim, $12 - 14$ magnitude, group.

For simplicity, we focus on a magnitude 11 star and magnitude 13 star to represent the two groups, and use the simulator to generate the relevant a's and then c's. Figure 6.7 shows the distribution of a's over the impact and timing parameters at opposition in a magnitude 13 star for four types of occultations; diameters 2km and 5km, distances 43AU and 300AU. Approximating the distribution of a over magnitudes $12 - 14$ with the distribution at magnitude 13, we take products with the signal to noise ratios to get the c's, and hence the power in figure 6.8. It is worth pointing out that of the histogram of c values in the top of figure 6.7, the only values that matter are those in the middle panel, because the power is either 0 or 1 outside that region.

In figure 6.9, we use the argument above to calculate the power of detection for objects in the range $[0.5km, 10km]$ for magnitudes 11 and 13, for distances $43AU$ and $300AU$, and for elongations 0° and 70°.

We recall from from chapter 3 that the *weighing factor* for an object of diameter D at distance R is given by

$$f_i(D, R) = \frac{E * v^* * H}{R^2}$$

where E is the exposure time, v^* is the relative velocity perpendicular to the line of sight, and H is the event cross section, which is increasing in D, R, the Fresnel scale, and the angular size of the star, θ_*. We note:

- The power for a magnitude 11 star ($M11$) is better than $M13$ for all comparisons considered. This means that given an object uniformly distributed over the range of detectability of the star, we have greater probability of detection in the $M11$ star. The $M11$ also has a bigger range of detectability: bigger magnitude means bigger θ_* which means bigger weighting factor when all else is held equal. For this Gaussian

single point model, $M11$ stars are more valuable for the survey efficiency than $M13$ stars.

- At elongation 70°, there is slightly higher power than at opposition for all the examples considered. However, since the relative velocity is smaller, those would cover less effective area. Since the weighting factor is linear in v^*, the appropriate adjustment to the probabilities would be the ratio of relative velocities. Assuming circular orbits at $43AU$, the relative velocities are $25.25km/s$ and $14.3km/s$, giving the ratio 1.76. For objects at $43AU$ of diameter 2km-3km and smaller, it appears that observing at elongation 70° will lead to a greater expected number of events than opposition. For the other cases, the probability of detection is not 76% higher at the elongation of 70°, so the expected number of events will be higher at opposition. At $300AU$ the ratio is 1.62, and there may be a slight advantage for 3km in M11 stars elongation 70°, but otherwise opposition is better.

- It turns out that most TAOS observations are taken near opposition, and for a randomly chosen $M13$ stars at opposition, the probability of detecting anything diameter $2km$ or smaller is less than 2%. For $M11$ stars and objects at $43AU$, the detection probability is 20%

In figure 6.10 we see the impact of a 10% noise reduction for our four types of signals. The improvement for magnitude 11 stars is less than a percent or two. For magnitude 13 stars, the advantage is a little better but still not great. This argument based on the normal distribution gives some theoretical support to the fact that our attempts to improve power using regression did not work very well.

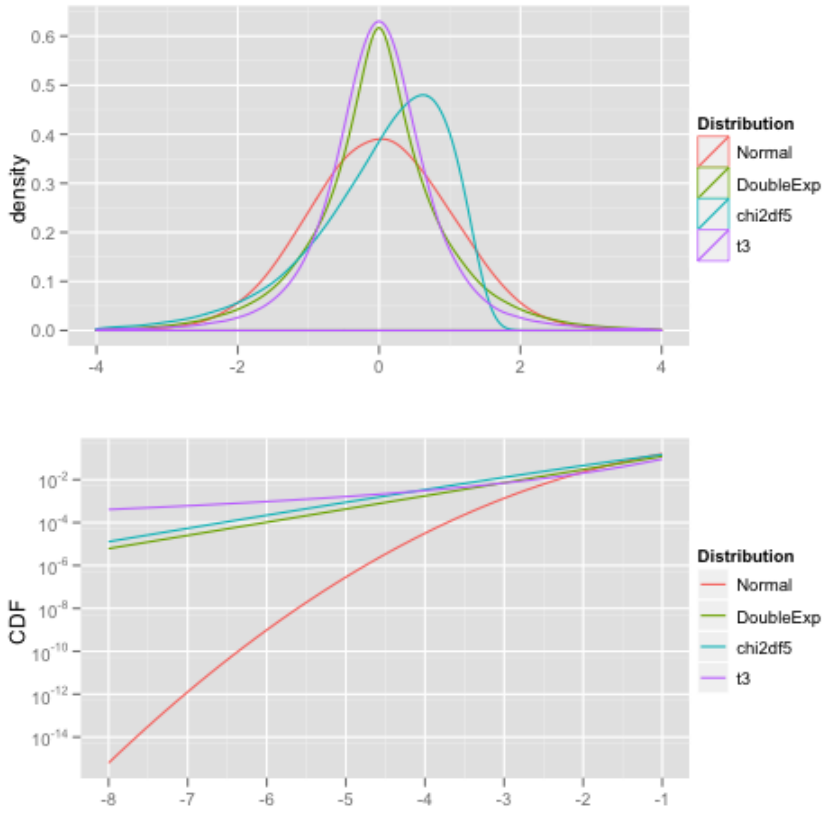

Figure 6.1: **Top:** Densities for the Normal distribution and centered and scaled versions of double exponential distribution, t distribution with 3 degrees of freedom, and negative of a chi square distribution with 5 degrees of freedom. **Bottom:** Lower left hand tails of the centered and scaled distributions.

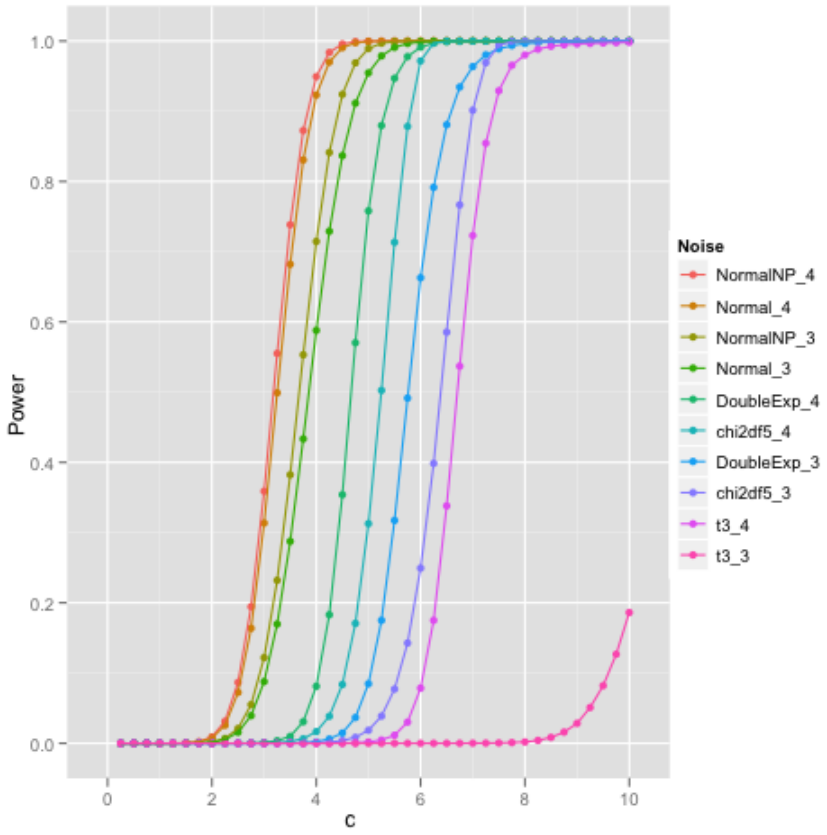

Figure 6.2: Power curves for various error terms, for the three and four telescope case.

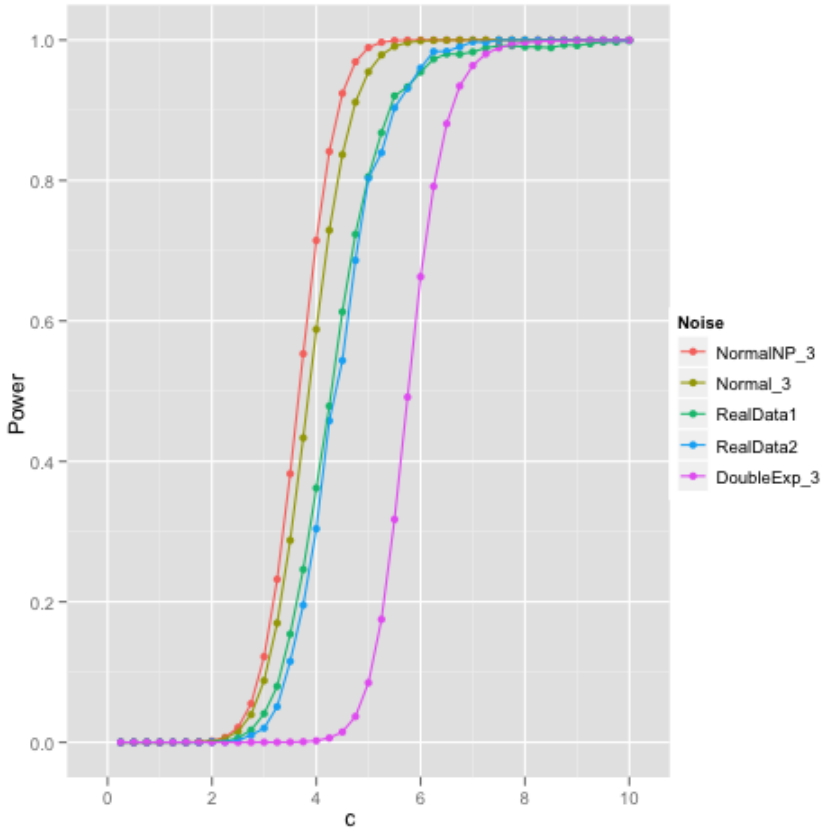

Figure 6.3: Power curves for some three telescope data along with RealData, representing the experiment in chapter 5

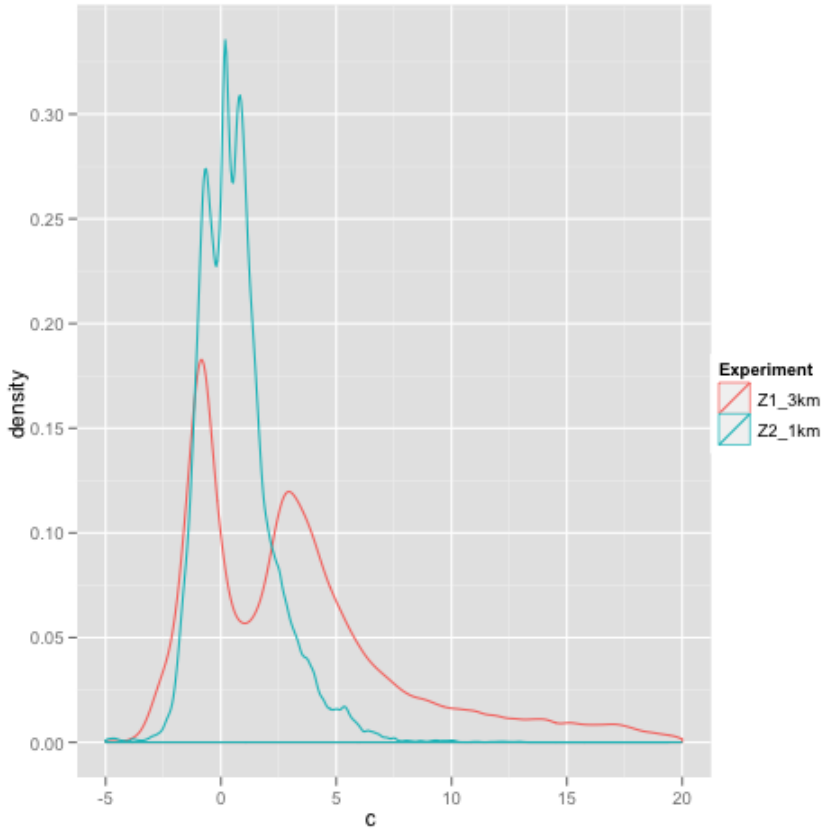

Figure 6.4: Distribution of the parameter $c = a * \frac{\mu}{\sigma}$ for the two experiments, Zipper Run #1 with $3km$ object and Zipper Run #2 with $1km$ object. The density plot is truncated to show $(-5, 20)$, although both distributions extend below -9 and the $3km$ run has some points as high as 35. However, based on figure 6.3, we might as well set everything less than 1 to 1 and everything greater than 10 to 10.

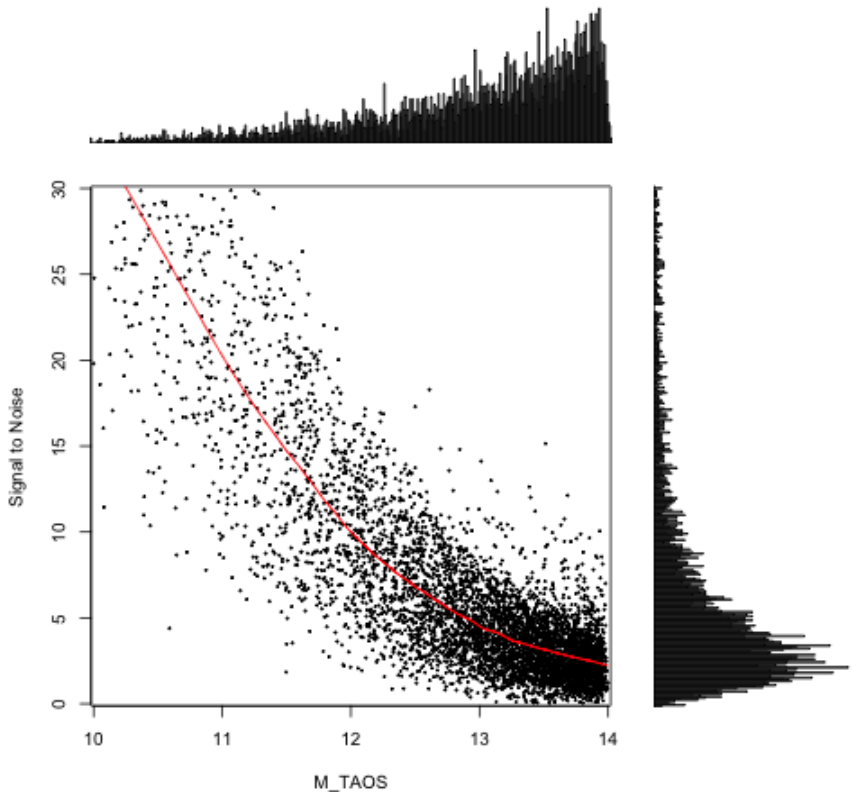

Figure 6.5: Signal to noise ratio vs. magnitude for a sample of TAOS stars. The signal to noise ratio is the average over the light curve triple. The points come from a random sample of 1% of the light curve triples. This plot is truncated at magnitude 10 and signal to noise 30, so that 4% of the data is not shown.

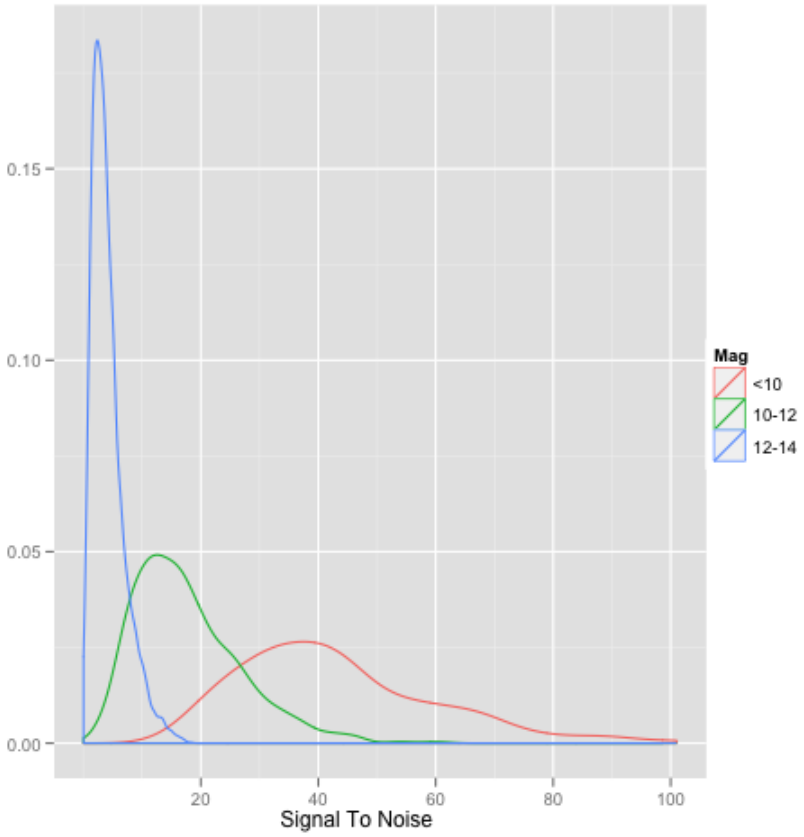

Figure 6.6: Density plots of signal to noise ratios for three groups based on magnitude. The blue group represent about 83% of the data, the green group is about 15% of the data, and the red group is the remaining 2%.

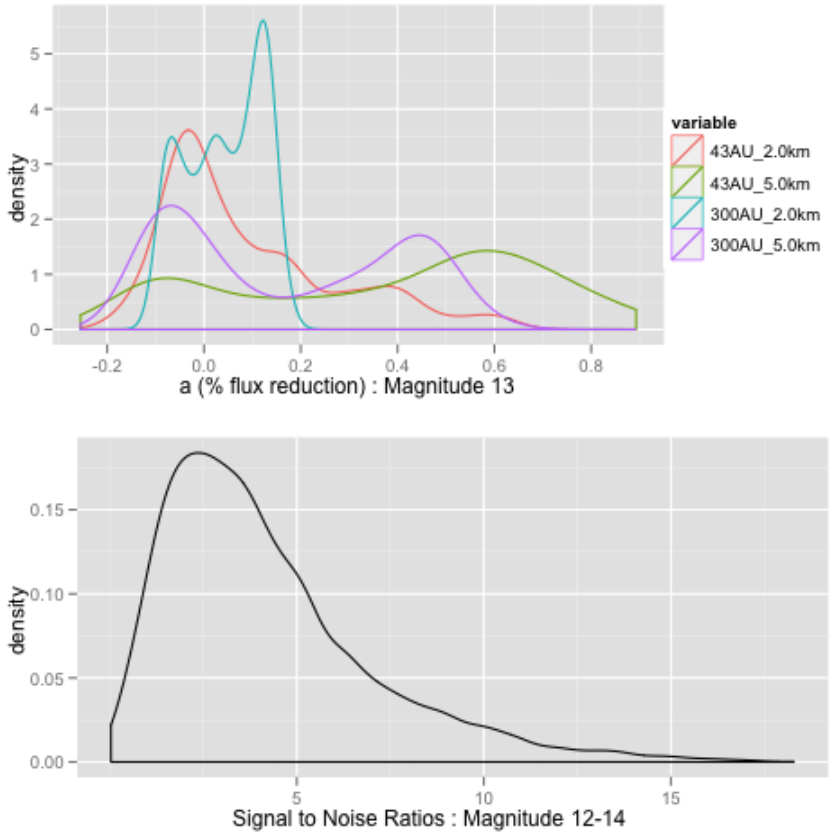

Figure 6.7: Distributions of % flux reductions at opposition and signal to noise ratios for dim stars.

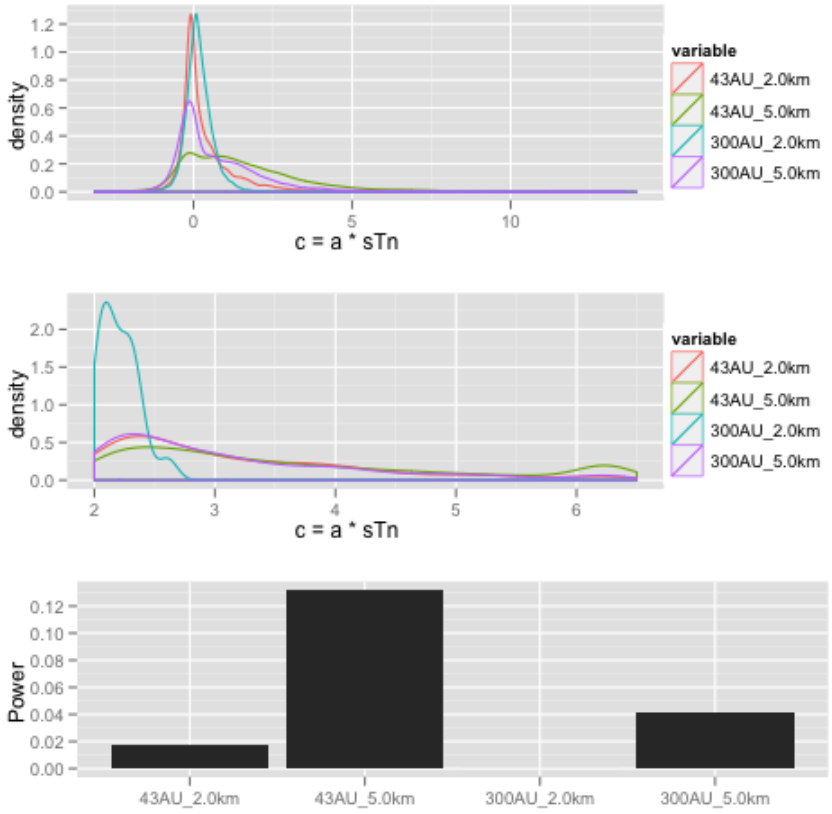

Figure 6.8: **Top Plot:** product of distributions in figure 6.7 for the four types of signals. **Middle Plot:** The top plot is only shown for $c > 2$ because there is no power when $c < 2$. Also, all values $c > 6$ are set to $c = 6.25$, because all such values have almost perfect power. **Bottom Plot:** Power to detect each of the four types of occultations. This is an integral of the density in the middle plot against the Gaussian power in figure 6.3.

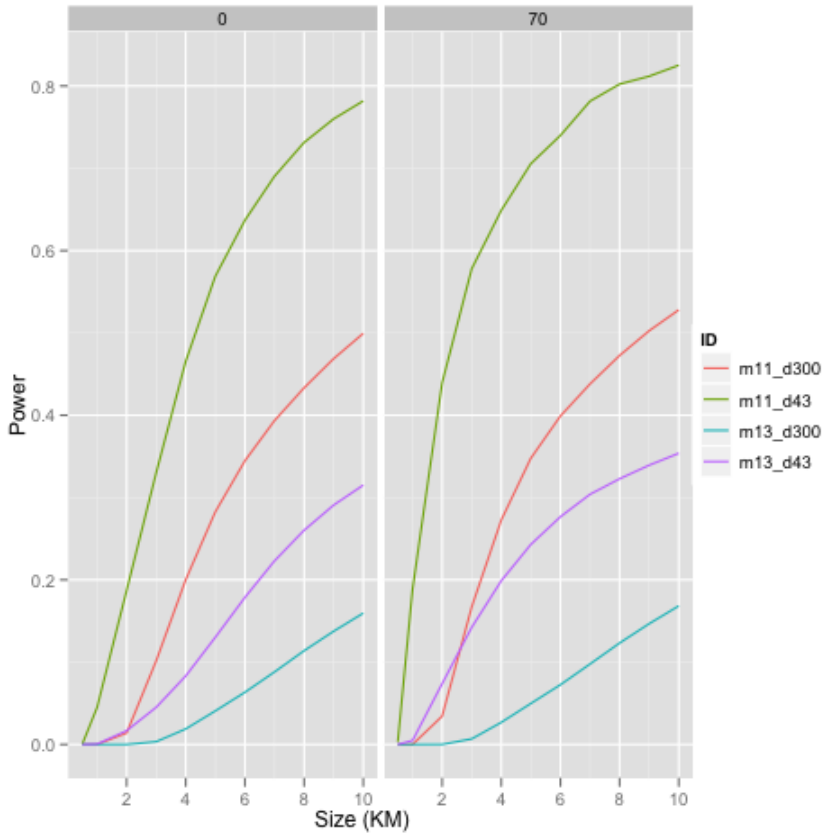

Figure 6.9: **Left Hang Panel:** Power Vs Size at opposition where we are averaging uniformly over the distribution of signal to noise ratio's for the given magnitude of star. **Left Hang Panel:** Same plot but at elongation 70°

.

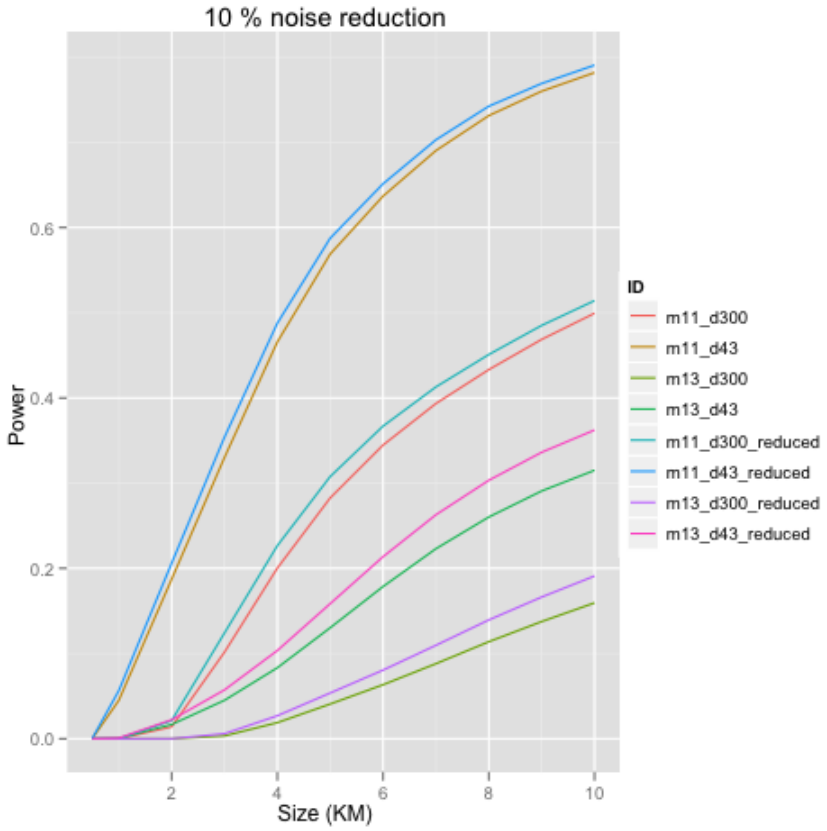

Figure 6.10: The impact of a 10% noise reduction at opposition.

Chapter 7

Fisher's Method

7.1 Optimal Properties of Fisher's Method

Given p-values from independent hypothesis tests, Fisher [Mosteller and Fisher, 1948, Fisher, 1956] proposed testing the global null hypothesis by using the product of p-values $T = \prod_{i=1}^{K} p_i$. As noted in Chapter 2,

$$-2log(T) = \sum_{k=1}^{K} -2log(p_k) \stackrel{d}{=} \sum_{k=1}^{K} \chi_2^2 \stackrel{d}{=} \chi_{2K}^2$$

In this chapter we look at situations where Fisher's method is optimal.

Lemma 7.1.1. *Likelihood Ratio and Invariance I*

For a simple hypothesis test with continuous densities, the likelihood ratio is equal to $h(p)$, the density of p-values under the alternative:

$$LR(x) \propto h(p) = f_{\theta_A}(F_{\theta_N}^{-1}(p))|(F_{\theta_N}^{-1})'(p)|$$

In addition, if the densities are differentiable, then so is $h(p)$.

Proof. Suppose $x \sim F_\theta \quad \text{for} \quad \theta \in \{\theta_N, \theta_A\}$ and let f_θ be the density of F_θ.

Consider the likelihood ratio:

$$LR(x) = \frac{f_{\theta_A}(x)}{f_{\theta_N}(x)}$$

Now let g be strictly increasing and differentiable and let $Y = g(X)$, then

$$f_Y(y) = f_\theta(g^{-1}(y))|(g^{-1})'(y)|$$

Therefore

$$LR(y) = \frac{f_{\theta_A}(g^{-1}(y))|(g^{-1})'(y)|}{f_{\theta_N}(g^{-1}(y))|(g^{-1})'(y)|} = \frac{f_{\theta_A}(g^{-1}(y))}{f_{\theta_N}(g^{-1}(y))} = \frac{f_{\theta_A}(g^{-1}(g(x)))}{f_{\theta_N}(g^{-1}(g(x)))} = LR(x)$$

So the likelihood ratio is invariant to such transformations. In particular, if we let $g = F_{\theta_N}$ then the denominator becomes uniform and our result follows. □

If we have K hypothesis that are independent under the null, then we can transform as above using the Jacobian, and note that the denominator is still unity:

Lemma 7.1.2. *Likelihood Ratio and Invariance II*

For K simple hypothesis test with continuous densities that are independent under the null hypothesis, the likelihood ratio is equal to $h(p_1, ..., p_K)$, the density of p-values under the alternative. If the densities under the null and alternative are smooth, than h is also smooth.

We note that Fisher's method rejects the Null for large values of $f(p_1, ..., p_K) = \prod_{k=1}^{K} \frac{1}{p_k}$. On the other hand, if we have independence under the Null, then the Neyman-Pearson lemma rejects for large values of $h(p_1, ..., p_K)$. This implies that if $h = \lambda \circ f$ for some increasing function λ, then Fisher's method is optimal. The converse is also true:

Theorem 7.1.3. *Optimality of Fisher's Method*

Suppose there are K simple hypothesis tests that are independent under the Null.

$$x_k \sim F_{\theta_k} \quad for \quad \theta \in \{\theta_{(N,k)}, \theta_{(A,k)}\}$$

Assume the densities under the null and the alternative are smooth. Then Fisher's method is optimal ***if and only if*** *there exists an increasing, continuous function λ such that*

$$h(p_1, ..., p_K) = \lambda(\prod_{k=1}^{K} \frac{1}{p_k})$$

Proof. Of Equivalence

Fisher's method is optimal **if and only if** there exists a function $\lambda(c)$ such that

$$\{p : f(p) \geq c\} = \{p : h(p) \geq \lambda(c)\}$$

Or, equivalently

$$f^{-1}[c, \infty) = h^{-1}[\lambda(c), \infty)$$

We let μ denote lebesgue measure on $[0, 1]^K$, and proceed in three steps.

1) $\lambda(x)$ must be strictly increasing

In general we have

$$f^{-1}[c + \epsilon, \infty) \subseteq f^{-1}[c, \infty)$$

And since $\mu(f^{-1}[c, c+\epsilon)) > 0$ we have strict inclusion

$$f^{-1}[c+\epsilon, \infty) \subset f^{-1}[c, \infty)$$

So,

$$h^{-1}[\lambda(c+\epsilon), \infty) \subset h^{-1}[\lambda(c), \infty)$$

Which implies $\lambda(c+\epsilon) > \lambda(c)$

2) $\lambda(x)$ **must be continuous**

Here we use the fact that discontinuities of monotonic functions must be 'of the first kind', meaning that the left and right limits always exist. For example, see Theorem 4.29 in [Rudin, 1976].

Suppose $\lambda(x)$ is discontinuous at c, and without loss of generality assume the limit from above is $\lambda(c) + a$ for some $a > 0$.

Note that

$$\begin{aligned} h^{-1}[\lambda(c), \lambda(c+\epsilon)) &= h^{-1}[\lambda(c), \infty) \cap (h^{-1}[\lambda(c+\epsilon), \infty))^c \\ &= f^{-1}[\lambda(c), \infty) \cap (f^{-1}[c+\epsilon, \infty))^c \\ &= f^{-1}[c, c+\epsilon) \end{aligned}$$

As $\epsilon \to 0$ from above, we have $\mu(f^{-1}[c, c+\epsilon)) \to 0$, which implies that

$$\mu(h^{-1}[\lambda(c), \lambda(c)+a)) = 0$$

Since we assume the densities were smooth, h must be continuous. Therefore, $\lambda(c)$ must be above the range that h can attain so it is not of interest. This implies that $\lambda(x)$ must be continuous for the relevant x.

3) $\lambda \circ f = h$

Since λ is strictly increasing and continuous, it has a continuous inverse λ^{-1}. Therefore,

$$f^{-1}[c, \infty) = (\lambda^{-1} \circ h)^{-1}[c, \infty)$$

Where $T(x) = (\lambda^{-1} \circ h)(x)$ must be continuous, being the composition of two continuous functions. Since any closed set, including the singleton $\{c\}$, can be obtained from topological operations on the closed intervals $[c, \infty)$, this gives us equality of functions:

$$f = \lambda^{-1} \circ h$$

And the result follows.

□

This tells us that if λ is an increasing, continuous function and $\lambda(\prod_{k=1}^{K} \frac{1}{p_i})$ integrates to one, then Fisher's method is optimal for $h(p) = \lambda(\prod_{k=1}^{K} \frac{1}{p_i})$.

Corollary 7.1.4. *Mixtures*
Suppose we have independence under the Null, and suppose the alternative is a mixture distribution

$$f_A(x_1, ...x_K) = \sum_{i=1}^{M} q_i f_A(x; \theta_i)$$

So that,

$$\begin{aligned} h(p) &= LR(x) \\ &= \frac{\sum_{i=1}^{M} q_i f_A(x; \theta_i)}{f_N(x)} \\ &= \sum_{i=1}^{M} q_i LR(x; \theta_i) \\ &= \sum_{i=1}^{M} q_i h(p; \theta_i) \end{aligned}$$

If Fisher's method is optimal for each $h(p; \theta_i)$, then Fisher's method is also optimal for $h(p)$.

Proof. This is true since a convex combinations of increasing, continuous functions is increasing and continuous. □

A concrete example is:

Example. *Power Densities*

First we note that if $b \in (-1, 0)$ then $\int_0^1 (b+1)p^b dp = 1$.

This implies that a density is given by

$$g(p; b) = (b+1)^K \left(\prod_{k=1}^{K} \frac{1}{p_k} \right)^{-b}$$

So for any mixture of $b_1, ..., b_M \in (-1, 0)$ with probabilities $q_1, ..., q_M$, Fisher's Method is optimal for

$$\begin{aligned} g(p) &= \sum_{i=1}^{M} q_i g(p|b_i) \\ &= \sum_{i=1}^{M} q_i (b_i+1)^K \left(\prod_{k=1}^{K} \frac{1}{p_k} \right)^{-b_i} \end{aligned}$$

A special case to consider is where we have independence under the alternative.

Theorem 7.1.5. *Optimality of Fisher's Method II*

Suppose that in addition to the conditions listed in Theorem 7.1.3, we have independence in the alternative. Then Fisher's method is optimal ***if and only if*** $\exists b \in (-1, 0)$ *such that*

$$h_k(p_k) = (b+1)p_k^b \quad \forall k$$

Where $h_k(p_k)$ *is the density of p values under the alternative for hypothesis k.*

Corollary 7.1.6. *Under the conditions of Theorem 7.1.5,*
Fisher's method is optimal ***if and only if*** $\exists \gamma \in (0, 1)$ *such that for each individual hypotheses the power-level relationship is given by*

$$Power(\alpha) = \alpha^\gamma$$

Proof. To get this result, integrate the density of p-values under the alternative and substitute $b + 1 \to \gamma$ □

Proof. Of Theorem 7.1.5
One direction is trivial, based on the example above.

For the other direction, note that with independence under the alternative, the distribution of the p-values under the alternative factorize:

$$\begin{aligned} h(p_1, ..., p_K) &= LR(x_1, ..., x_K) \\ &= \frac{f_A(x_1, ..., x_K)}{f_N(x_1, ..., x_K)} \\ &= \prod_{k=1}^{K} \frac{f_A(x_k)}{f_N(x_k)} \\ &= \prod_{k=1}^{K} LR(x_k) \\ &= \prod_{k=1}^{K} h_k(p_k) \end{aligned}$$

Giving us

$$\lambda \circ f = \prod_{k=1}^{K} h_k(p_k)$$

Since the left hand side is invariant to permutations of its arguments, the right hand side must be, so $h_i = h_j \quad \forall i, j$, which implies

$$\lambda(\prod_{k=1}^{K}\frac{1}{p_k}) = \prod_{k=1}^{K} h_1(p_k)$$

We define $h = \prod_{k=1}^{K} h_1(p_k)$ and take partial derivatives of both sides of $\lambda \circ f = h$

$$\begin{aligned}\frac{\partial}{\partial p_k}\Big((\lambda \circ f)(p)\Big) &= \lambda'(f(p))\frac{\partial}{\partial p_k}f(p) \\ &= \lambda'(f(p)) * f(p) * (-\frac{1}{p_k})\end{aligned}$$

And

$$\frac{\partial}{\partial p_k}h(p) = h(p) * \frac{h_1^{'}(p_k)}{h_1(p_k)}$$

Setting these expressions equal and then isolating the terms depending on k gives

$$-\lambda'(f(p))\frac{f(p)}{h(p)} = \frac{p_k h_1^{'}(p_k)}{h_1(p_k)}$$

Since the equality above must be true for all k, the left hand side must be constant in p. This implies that there exists a b such that

$$\frac{p_k h_1^{'}(p_k)}{h_1(p_k)} = b \qquad \forall k$$

This is the differential equation

$$\frac{1}{h}dh = \frac{b}{p}dp$$

And integration gives the unique family of solutions $h_1(p) = c * p^b$. Since h_1 is a density, $b \in (-1, 0)$, and $c = b + 1$.

□

Chapter 8

Multi Point Theory

In this chapter we consider parametric models for occultations lasting more than one time point. We derive the optimal test statistic in some cases, consider some practical sub-optimal methods, and mention their relationship to methods in the literature. We also discuss how one might estimate size and distance in the presence of an occultation.

8.1 Setup and Notation

We define:

$$\begin{aligned} n &= \text{The length of the signals of interest} \\ \lambda &= \frac{\mu}{\sigma} = \text{The signal to noise of the light curve} \\ I(\theta) &= \text{a defraction profile with baseline 1 and parameter } \theta \\ C(\theta) &= I(\theta) - \mathbf{1} = \text{the defraction profile with baseline 0} \\ \theta &= \text{parameter (timing, impact, size, distance)} \end{aligned}$$

Where our parameter space is given by

$$\Theta \subset (0, 95ms) \times (0, H/2) \times A \qquad \text{With} \qquad A \subset (0, 1000km) \times (30AU, 10000AU)$$

We will sometimes assume θ has a prior distribution over $\Theta = \{\theta_1, ..., \theta_M\}$ with probabilities $p_1, ..., p_M$.

8.2 Optimal Detection

We consider the signal detection problem in the Poisson and Normal cases. If we are looking for one known signal, or for many signals indexed by a prior distribution, then the Neyman-Pearson lemma tells us the optimal method. Here we find the optimal tests in those cases, and note that K independent telescopes can be reduced to the one telescope in both models.

8.2.1 The Poisson Model

Assume that, conditional on θ, our data is generated independently on K telescopes according to

$$z_i^k \sim \text{Poisson}(\mu * I_i(\theta)) \qquad 1 \leq i \leq n \qquad 1 \leq k \leq K$$

Where we assume that μ is a known constant and we allow for a prior distribution over θ. Then the likelihood is given by

$$\begin{aligned}
f(z|\theta) &= \prod_{k=1}^{K}\prod_{i=1}^{n} \frac{\exp(-\mu I_i)(\mu I_i)^{z_i^k}}{z_i^k!} \\
&= \prod_{k=1}^{K}\prod_{i=1}^{n} \exp\Big(-\mu I_i + z_i^k log(\mu I_i) - log(z_i^k!)\Big) \\
&= \exp\Big(\sum_{k=1}^{K}\sum_{i=1}^{n} \big(-\mu I_i + z_i^k log(\mu I_i) - log(z_i^k!)\Big) \\
&= \exp\Big(-K\mu \sum_{i=1}^{n} I_i + \sum_{i=1}^{n} log(\mu I_i) \sum_{k=1}^{K} z_i^k - \sum_{i,k} log(z_i^k!)\Big) \\
&= \exp\Big(-K\mu n\bar{I} + \sum_{i=1}^{n} log(\mu I_i) z_i^* - \sum_{i,k} log(z_i^k!)\Big) \\
&= \exp\Big(-K\mu n - \sum_{i,k} log(z_i^k!) + \sum_{i=1}^{n} log(\mu) z_i^* - K\mu n(\bar{I}-1) + \sum_{i=1}^{n} log(I_i) z_i^*\Big) \\
&= f(z|N)\exp\Big(-Kn\mu(\bar{I}-1)\Big)\exp\Big(< log(I), z^* > \Big)
\end{aligned}$$

Where $f(z|N)$ denotes the likelihood under the null that $I = 1$, and $log(I)$ means we take the log point-wise, $\bar{I} = \frac{1}{n}\sum_{i=1}^{n} I_i$, and we introduced the sufficient statistic $z_i^* = z_i^1 + z_i^2 + ... + z_i^K$. This implies that our likelihood ratio given a prior over templates is

$$LR(z) = LR(z^*) = \sum_{i=1}^{M} p_i \exp\Big(Kn\mu(\bar{I}(\theta_i) - 1)\Big)\exp\Big(< log(I(\theta_i)), z^* > \Big)$$

Hence, given K independent telescopes in the Poisson model, the optimal way to combine the data is to add them, which corresponds to one telescope data with mean $K * \mu$.

8.2.2 The Normal Model

We assume our data is generated independently on the K telescopes according to

$$y_i^k = \mu_k I(t_i;\theta) + \sigma_{i,k}\epsilon_{i,k} \qquad \text{for} \quad i = 1, ..., n \qquad k = 1, ..., K, \quad \text{With} \quad \epsilon_{i,k} \overset{iid}{\sim} \mathcal{N}(0,1)$$

Where $\mu_k, \sigma_{i,k}$ are known constants and we allow for a prior over θ. We change variables for convenience and set

$$z_i^k = \frac{y_i - \mu_k}{\mu_k} = (I(t_i;\theta) - 1) + \frac{\sigma_{i,k}}{\mu_k}\epsilon_{i,k}$$

Then we put it in vector notation:

$$z^k = C(\theta) + D_k \epsilon^k$$

We also have interest in the special case of homoscedastic errors on each telescope, so that D_k only depends on the signal to noise: $D_k = \frac{1}{\lambda_k} I$

We have

$$\begin{aligned}
f(z|\theta) &= \prod_{k=1}^{K} f(z^k|\theta) \\
&= \prod_{k=1}^{K} (\frac{1}{2\pi})^{n/2} |D_k| \exp\left(-\frac{1}{2}(z^k - c(\theta))' D_k^{-2} (z^k - c(\theta))\right) \\
&= (\frac{\lambda_k}{\sqrt{2\pi}})^{\frac{nK}{2}} (\prod_{k=1}^{K} |D_k|)^{-1} \exp\left[-\frac{1}{2}\sum_{k=1}^{K}(z^k - c(\theta))' D_k^{-2}(z^k - c(\theta))\right] \\
&= f(z|N) * \exp\left[-\frac{1}{2}\sum_{k=1}^{K} C(\theta)' D_k^{-2} C(\theta)\right] * \exp\left[\sum_{k=1}^{K} < D_k^{-2} z, C(\theta) > \right] \\
&= f(z|N) * \exp\left[-\frac{1}{2} C(\theta)' D_*^{-2} C(\theta)\right] * \exp\left[< D_*^{-2} z^*, C(\theta) > \right]
\end{aligned}$$

In the fourth equality, $f(z|N)$ denotes the likelihood under the null that $C(\theta) = 0$. In the fifth equality we introduced $D_* = (\sum_{k=1}^{K} D_k^{-2})^{-1/2}$, so it would satisfy $(\sum_{k=1}^{K} D_k^{-2}) = D_*^{-2}$. We also introduced the sufficient statistic

$$z^* = D_*^2 \sum_{k=1}^{K} D_k^{-2} z^k$$

In the special case of $D_k = \frac{1}{\lambda_k} I$, this becomes $z^* = \frac{1}{\lambda^*}\sum_{k=1}^{K} \lambda_k^2 z^k$, where the new signal to noise ratio is $\lambda_* = \sqrt{\sum_{k=1}^{K} \lambda_k^2}$.

This implies that our likelihood ratio given a prior over templates is

$$LR(z) = LR(z^*) = \sum_{i=1}^{M} p_i \exp\left[\frac{-1}{2}(C(\theta_i))' D_*^{-2} (C(\theta_i))\right] \exp\left[< D_*^{-2} z^*, C(\theta) > \right]$$

Hence, given K independent telescopes in the Normal homoscedastic model, the optimal way to combine the data is through the sufficient statistic z^*, which is equivalent to

a single telescope model with $D = D_*$. In the special case where all telescopes have constant errors and the same signal to noise ratio, the combination is equivalent to the signal to noise ratio increasing by a factor if $\sqrt{K}$. From here forward, we restrict our attention to the one telescope case.

8.2.3 Computation and Distribution

The distribution of the likelihood ratio is extremely skewed, but the log likelihood ratio, $LLR(z) = log(LR(z))$ has a more natural shape. The log likelihood ratio can be written in terms of the *LogSumExp* function where

Definition 8.2.1. *LogSumExp*
If v_i is the i'th element of the M-vector v, then we write

$$LogSumExp\Big(v\Big) = log\Big(\sum_{i=1}^{M} \exp(v_i)\Big)$$

LogSumExp is a convex function that is related to the max. The relationship to the max helps with numerical stability issues. Properties and numerical issues related to the *LogSumExp* are illustrated in the Appendix section 10.2.

Model		Argument to LogSumExp (i'th component)
Poisson	μ	$log(p_i) + n\mu\bar{C}(\theta_i) + < z, log\big(C(\theta_i)+1\big) >$
Normal	D	$log(p_i) + \frac{-1}{2}C(\theta_i)'D^{-2}C(\theta_i) + < D^{-2}z, C(\theta_i) >$
Normal	$\frac{1}{\lambda}I$	$log(p_i) + \frac{-\lambda^2}{2}C(\theta_i)'C(\theta_i) + \lambda^2 < z, C(\theta_i) >$

Since the argument to *LogSumExp* is affine in z, and the function *LogSumExp* is convex, this tells us that $LLR(x)$ is convex. We note that

- If we are searching for one signal I in the Poisson model, we reject the null for large values of $< z, log(I) >$.
- If we are searching for one signal I in the Normal model, we reject the null for large values of $< z, I >$.
- If we are searching for several signals indexed by a prior in either model, then our acceptance region is a level set of $LLR(z)$, which must be a convex set.

Given a collection of signals indexed by a prior, we can find level α thresholds for $LLR(z)$ using importance sampling. This is explained in detail in the Appendix section 10.3.

8.3 Sub-Optimal Methods

Here we mostly put ourselves in the Normal homoscedastic model and consider the power of detection methods based on some simple test statistics. Based on the arguments above, we restrict ourself to the 1 telescope case. We consider power for individual signals and for families of signals indexed by a prior.

8.3.1 Midpoint Reduction

If we let a denote the flux reduction at the midpoint,

$$a(\theta) = -1 * \text{Midpoint of } C(\theta)$$

Then, we saw in Chapter 6 that the power of detection is

$$POW(\theta, \lambda, \alpha) = \Phi(\Phi^{-1}(\alpha) + \lambda a(\theta))$$

The power over a collection of signals with a prior p_i is

$$POW(\alpha) = \sum_{i=1}^{M} p_i \Phi\left(\Phi^{-1}(\alpha) + c_i\right)$$

Where $c_i = \lambda * a(\theta_i)$.

We could also take a moving average of J points before applying the procedure. In the i.i.d. case, the new value of c is then given by

$$\begin{aligned} c(J) &= a(J)\lambda(J) \\ &= a(J)\frac{1}{\sqrt{J}}\lambda \\ &= c(1) * \left(\frac{1}{\sqrt{J}}\frac{a(J)}{a(1)}\right) \end{aligned}$$

8.3.2 The Norm Test Statistic

Since occultation patterns produce flux increases and decreases for several time points, a possible test statistic is the sum of squares of deviations, or the *Norm* of the data. This test and its power depend on the central and non-central χ^2 distributions.

Under the null,

$$||z||^2 = (\frac{1}{\lambda})^2||\epsilon||^2 \stackrel{d}{=} (\frac{1}{\lambda})^2\chi^2_n$$

We let $F_{\chi^2_n}$ denote the upper tail of χ^2_n, and $F_{\chi^2_{n,a}}$ denote the upper tail of the non-central chi-square distribution with non-centrality parameter a. Then p-values are given by

$$p = F_{\chi^2_n}(||\lambda z||^2)$$

Under the alternative this norm follows a non-central chi-square distribution,

$$||\lambda z||^2 \sim \chi^2_{n,c^2} \qquad \text{With} \qquad c = \lambda||C(\theta)||$$

This gives us

$$\begin{aligned} POW(c(\theta), \alpha, n) &= POW(C(\theta), \lambda, \alpha, n) \\ &= \mathbb{P}(\chi^2_{n,c^2(\theta)} > F^{-1}_{\chi^2_n}(\alpha)) \\ &= F_{\chi^2_{n,c^2(\theta)}}(F^{-1}_{\chi^2_n}(\alpha)) \end{aligned}$$

This function of (n, c) can be seen in figure 8.1 for level $\alpha = 10^{-10}$.

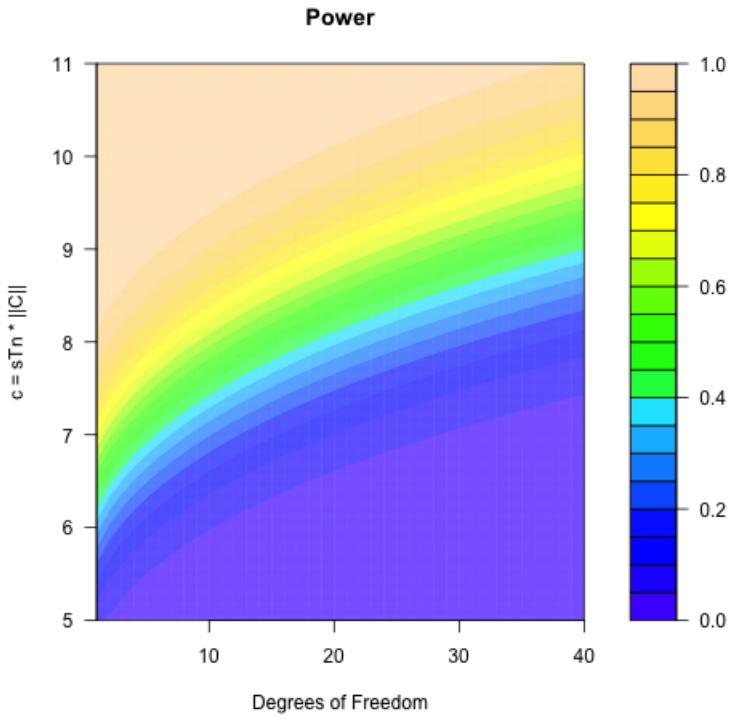

Figure 8.1: Power of the chi square test in terms of d.f. and parameter $c = \lambda * ||C||$

According to the contours in figure 8.1, when degrees of freedom increase, the value of c must also increase in order to maintain the same level of power. Intuition based on the normal approximation is that increasing the degrees of freedom increases the variance while the deviation from the mean stays fixed, which shrinks the corresponding z score.

Given a collection of signals with prior according to p_i, we have

$$POW(\alpha) = \sum_{i=1}^{M} p_i F_{\chi^2_{n,c_i^2}}(F^{-1}_{\chi^2_n}(\alpha))$$

Where $\quad c_i = \lambda||C(\theta_i)||$

8.3.3 Linear Regression

Let X be an n by p matrix, with an intercept, designed such that for each θ, there exists a β such that

$$C(\theta) \approx X\beta$$

This parameterization is nice because $\beta = 0$ corresponds to our null hypothesis. Before looking at methods, we recall a couple facts about linear models.

Linear Models and the F-Test

Consider

$$Y = X\beta + \epsilon \qquad \text{Where} \quad \epsilon \sim \mathcal{N}(0, \sigma^2 I)$$

Here X is n by p, and contains an intercept. We will compare the fit using all the variables to the fit based on the intercept. We let $\hat{\beta}$ denote the OLS estimate, e denote the residual, and the superscript s will denote for the small model (intercept only). Then standard theory (see [Freedman, 2005], for example) tells us

1. $$\frac{||e^s||^2 - ||e||^2}{\sigma^2} \sim \chi^2_{p-1}$$
2. $$\frac{||e||^2}{\sigma^2} \sim \chi^2_{n-p}$$
3. $$\frac{n-p}{p-1}\frac{||e^s||^2 - ||e||^2}{||e||^2} \sim F_{(p-1,n-p)}$$

Where rejections based on the F statistic in the third line are equivalent to the likelihood ratio test.

GLS extension and the $\Delta\chi^2$

If we put ourself in the more general setting with a diagonal covariance $D_{i,i} = \sigma_i$

$$Y = X\beta + D\epsilon \qquad \text{Where} \quad \epsilon \sim \mathcal{N}(0, I)$$

Multiplication by D^{-1} and using the results above gives

$$||e||^2 = \min_{\beta} \sum_{i=1} \big(\frac{Y_i - (X\beta)_i}{\sigma_i}\big)^2 \qquad ||e^{(s)}||^2 = \min_{\mu} \sum_{i=1} \big(\frac{Y_i - \mu}{\sigma_i}\big)^2$$

Astronomers often consider the $\Delta\chi^2$ statistic for significance levels:

$$\Delta\chi^2 = ||e||^2 - ||e^{(s)}||^2 \sim \chi^2_{p-1}$$

We note that in this model that it is the ratio $\frac{||e^{(s)}||^2}{||e||^2}$ rather than the difference $\Delta\chi^2$ that arises from the likelihood ratio.

The method used in [Schlichting et al., 2009]

For general function fitting, $\Delta\chi^2$ can be thought of as measuring the extent that the fit is better than the straight line fit

$$\Delta\chi^2 = \min_{\beta} \sum_{i=1} \left(\frac{Y_i - (f(\beta))_i}{\sigma_i}\right)^2 - \min_{\mu} \sum_{i=1} \left(\frac{Y_i - \mu}{\sigma_i}\right)^2$$

In the recent paper, [Schlichting et al., 2009] reports detection of one occultation from 40 Hz data collected on the Hubble space telescope. Their test statistic is the $\Delta\chi^2$, where the function used is not clearly stated and may not be linear. Theory gives the distribution of $\Delta\chi^2$ when the function is linear and the errors are Normal, and the distribution can be found in the general case using the bootstrap. In particular, in [Schlichting et al., 2009] they sample from the part of the light curve excluding the event, then find their value of $\Delta\chi^2$, then sample again, forming and empirical distribution of the statistic based on the empirical distribution of the noise. By taking a large number of samples, they are able to clearly see the tail of the distribution.

We note that the resampling method in [Schlichting et al., 2009] assumes the data are exchangeable, so it would not work with TAOS data because of autocorrelation remaining after the filter.

Uses with the homoscasdic Normal model

Based on the model

$$Z = X\beta + \frac{1}{\lambda}\epsilon$$

We can find thresholds for the F test or χ^2 test that are valid under the null

$$Z = \frac{1}{\lambda}\epsilon$$

Which is the same null as in the true model

$$Z = C(\theta) + \frac{1}{\lambda}\epsilon$$

The power of these tests can be easily assessed under simulation using:

$$||e||^2 = ||Z - X\hat{\beta}||^2 = ||(I - H)(C(\theta) + \frac{1}{\lambda}\epsilon)||^2$$

where $H = X(X'X)^{-1}X'$ is the hat matrix and

$$||e^s||^2 = ||Z - \bar{Z}||^2 = ||(C - \bar{C}) + \frac{1}{\lambda}(\epsilon - \bar{\epsilon})||^2$$

Simpler Regression Methods

A simpler method is ignore the origin and let X be an n by p matrix of orthogonal templates. Then

$$\hat{\beta} = X'z = X'C(\theta) + \frac{1}{\lambda}X'\epsilon \sim \mathcal{N}(X'C(\theta), \frac{1}{\lambda^2}I)$$

$$\lambda X'z \sim \mathcal{N}(\lambda X'C(\theta), I)$$

If we look at $T = ||\lambda X'z||^2$ rather than $||\lambda z||^2$ then the analysis is the same as Chi Square fitting except the degrees of freedom change and non-centrality parameter change:

$$(p, ||\lambda X'C(\theta)||^2) \quad \text{Instead Of} \quad (n, ||\lambda C(\theta)||^2)$$

This has the potential to do better than the regular chi square because it could capture most of the variation in the signals using less degrees of freedom; the potential benefits can be seen in figure 8.1. The power over a collection of signals with a prior p_i is computable in terms of the non-central chi-square distribution:

$$POW(\alpha) = \sum_{i=1}^{M} p_i F_{\chi^2_{p,b_i}}(F^{-1}_{\chi^2_p}(\alpha))$$

Where $\quad b_i = ||\lambda X'C(\theta_i)||^2$.

8.4 Sensitivity of Occultation Surveys

For the design and planning of future occultation surveys there are several parameters that influence the sensitivity of the survey. Some parameters include the distribution of stars and signal to noise ratios implied by the site and equipment, the number of stars that can be observed at once, the projected fraction of the day that is observing [1], and the regions of the outer solar system being probed.

If a proposed survey had Poisson or Normal errors, and we knew the prior of objects, then we would use the likelihood ratio above as our test statistic, and could then compute the probability of detection for each light curve. Then, based on the distribution of other survey properties, we could compute the probability of detection per unit time of the survey as a whole and compare proposed surveys based on that number.

Unfortunately, the errors will not be Poisson or Normal and we don't know the priors of objects; in fact, the 'prior', or distribution of those objects are the unknown quantity of scientific interest. Despite the problems with Normal errors, it is useful to see what Normal errors imply about survey sensitivity, beyond the single point exploration in Chapter 6.

Let the subscript or superscript h denote the number of observations per second in a proposed survey. We let $I_h(\theta)$ denote the diffraction profile for signal θ in an h-Hertz survey. We let $m(h)$ be the length of $I_h(\theta)$ and note $m(h) = h * m(1)$. If μ_1 is the average

[1] Ground based surveys can only observe at night, but space based surveys could have less down time.

number of photons per second from a star, then $\mu_h = \frac{1}{h}\mu_1$ is the average number of photons per time point in a h-Hertz survey [2]. We let σ_h be the noise level per observation in a h-Hertz survey. In the case of Poisson statistics, $\sigma_h = \sqrt{\mu_h} = \frac{1}{\sqrt{h}}\sigma_1$. In general, if we consider longer exposures to be the sum of independent shorter exposures we would get the same relation $\sigma_h = \frac{1}{\sqrt{h}}\sigma_1$ which implies the signal to noise relationship $\lambda_h = \frac{1}{\sqrt{h}}\lambda_1$.

While the relations for μ_h and $m(h)$ are clear, the relationship $\sigma_h = \frac{\sigma_1}{\sqrt{h}}$ might not be useful for making decisions: given a fixed budget, a survey designed to run at 20 Hz will have better signal to noise ratio than the result of binning every two points in a survey designed to run at 40 Hz. On the other hand, if we start with a survey running at 40 Hz and are curious what would happen if we binned the data, then the relation $\sqrt{2}\sigma_{40} = \sigma_{20}$ is reasonable.

We consider the model

$$y_h = \mu_h I_h(\theta) + \sigma_h \epsilon \qquad \text{With} \qquad \epsilon \sim \mathcal{N}(0, I_{m(h)\times m(h)})$$

Transforming as before gives

$$z_h = C_h(\theta) + \frac{\sigma_h}{\mu_h}\epsilon$$

8.4.1 The Norm Statistic and Survey Sensitivity

The norm statistic, $||z||^2$ can be used to analyze the sensitivity of a survey. There is one statistic to consider for all types of signals (although lengths will differ), and it does not require any assumptions about priors. In [Nihei et al., 2007] they consider the norm test statistic in different surveys, and look at something kind of like power of that test as a measure of sensitivity. We point out below that caution should be used in interpreting the norm statistics from surveys of different sampling rates.

We know from section 3.2 that the power of a level α test is given in terms of the non-central chi square distribution

$$POW(h, \theta, \alpha) = F_{\chi^2_{\left(m(h), a(\theta)\right)}}(F^{-1}_{\chi^2_{m(h)}}(\alpha))$$

Where

$$\begin{aligned} a(\theta) &= ||\lambda_h C_h(\theta)||^2 && (8.1)\\ &= (\frac{\mu_1}{\sigma_h})^2 \frac{1}{h}\left(\frac{1}{h}||C_h(\theta)||^2\right) && (8.2)\\ &= \mu_1^2 * \frac{1}{h * \sigma_h^2} * \Xi_h(\theta) && (8.3) \end{aligned}$$

Where we have introduced $\Xi_h(\theta) = \frac{1}{h}||C_h(\theta)||^2$, which [Nihei et al., 2007] calls the Detectability.

[2] Here I am ignoring the down time. Although TAOS operates at 5Hz, it's exposure time is 105ms, so it would really receive 10.5% of the photons from a 1 second exposure.

We note that $\Xi_h(\theta)$ is a Riemann sum and will converge to the integral as resolution (h) increases

$$\Xi_h(\theta) \longrightarrow \Xi(\theta) \stackrel{d}{=} \int (C(\theta)(t))^2 dt$$

Previously we noted that if longer exposures are considered the sum of independent shorter exposures, then $\sigma_1^2 = h\sigma_h^2$, which would imply

$$a(h, \theta) = \lambda_1^2 \Xi_(h)(\theta) \approx \lambda_1^2 \Xi(\theta)$$

where the approximation is for large h where the Riemann sum approximates the integral well.

So as h increases, $a(h)$ stays roughly constant, but the degrees of freedom $m(h) = h * m(1)$ increases. This implies that even the ideal case where $\sqrt{h}\sigma_h = \sigma_1$, **power will decrease with increasing resolution according to the norm test statistic** !

The reason for this undesirable property is that the norm statistic stays constant when moving to higher resolutions case, while the test statistic acts as if more noise is being added.

Sensitivity Calculation in [Nihei et al., 2007]

Given our test statistic $T = ||z||^2 = ||C(\theta) + \frac{1}{\lambda}\epsilon||^2$, an event is called **detectable** in [Nihei et al., 2007] for θ at level α if the probability (under the null) that it exceeds its expectation (under the alternative $A(\theta)$) is less than α.

$$\mathbb{P}_N(T > \mathbb{E}_{A(\theta)}(T)) < \alpha$$

Since $T \stackrel{d}{=} \frac{1}{\lambda^2}\chi_m^2$ under the null, this can be restated

$$\mathbb{E}(||\lambda C(\theta) + \epsilon||^2) \geq \chi_m^2(\alpha)$$

To put this in the setting of hypothesis testing, we could set up a level α test based on T by detecting events in $T \geq \frac{1}{\lambda^2}\chi_m^2(\alpha)$. Next, since $||\lambda C(\theta) + \epsilon||^2$ follows a non-central χ^2 distribution, it follows that **if the non-central χ^2 were symmetric**, then this criteria would be equivalent to saying 'Event θ is **detectable** if the power to detect it is greater than 1/2.'

For a central χ_m^2 distribution, the median and mean can be calculated explicitly and the median is less than the mean (as expected from the long right tail). For the non-central χ^2 distribution the median cannot be calculated explicitly, but it is shown in [Sen, 1989] that the median is still less than the mean [3]. Therefore, in terms of power, the definition of 'detectable' corresponds to 'Event θ is **detectable** if the power to detect it is greater than $c(\theta, h)$, where $c(\theta, h)$ is a number less than a 1/2 that depends on both θ and the sampling rate of the survey'.

In the high resolution case, if we assume $\sigma_1^2 = h\sigma_h^2$, then

[3] They also show that the median is sub-additive in the non-centrality parameter, as opposed to the mean which is well known to be additive

$$\begin{aligned}\mathbb{E}(||\lambda_h C(\theta)+\epsilon||^2) &= \lambda_h^2||C(\theta)||^2+m(h)\\ &= \lambda_1^2\Xi_h(\theta)+m(h)\end{aligned}$$

So an event is 'detectable' if $\lambda_1^2\Xi_h(\theta)+m(h)>\chi^2_m(\alpha)$, which is equivalent to

$$\Xi_h(\theta)*\frac{\lambda_1^2}{2m(h)}>\frac{\chi^2_m(\alpha)-m(h)}{2m(h)}\approx\epsilon\sim\mathcal{N}(0,1)$$

For large h, the left hand side will decrease with h, implying that the survey becomes less 'detectable' as resolution increases, which is not a desirable property of a comparative test statistic.

8.5 Estimation of θ

Suppose we are in the model

$$y_i=\mu*I(t_i;\theta)+\sigma_i\epsilon_i \quad \text{for} \quad i=1,...n, \quad \text{where} \quad \epsilon_i\overset{iid}{\sim}\mathcal{N}(0,1)$$

$$\text{Let} \quad z_i=\frac{y_i-\mu*I(t_i;\theta)}{\sigma_i} \quad and \quad \chi^2(\theta)=\sum_{i=1}^n z_i^2=\sum_{i=1}^n(\frac{y_i-f(t_i;\theta)}{\sigma_i})^2$$

The likelihood is

$$L(\theta|Y)=\mathbb{P}(Y|\theta)=\prod_{i=1}^n\mathbb{P}(z_i)\propto\exp(-1/2*\chi^2(\theta))$$

Hence, minimization of $\chi^2(\theta)$ is equivalent to maximum likelihood estimation. Since $I(t;\theta)$ may not be convex in θ, this could be a very difficult optimization problem with several local minima.

If we have a detection and solved the maximum likelihood optimization, we use it as our estimate of θ, but we would not have standard errors. Alternatively we could take a Bayesian perspective and put a prior on (t,i,s,d), the timing parameter, impact parameter, size of KBO, and distance to KBO. The posterier probability region would be based on

$$\begin{aligned}\mathbb{P}(t,i,s,d|Y) &= \mathbb{P}(Y|t,i,s,d)*\frac{\mathbb{P}(t,i,s,d)}{\mathbb{P}(Y)}\\ &\propto \chi^2(t,i,s,d)\mathbb{P}(t,i,s,d)\\ &\propto \chi^2(t,i,s,d)\mathbb{P}(s,d)\end{aligned}$$

Where the last step is assuming that timing and impact parameters are uniform and independent of the size and distance. Since the scientific interest is in size and distance, we would integrate out the other parameters, and inference would be based on

$$\mathbb{P}(s,d|Y) \propto \mathbb{P}(s,d) \int\int \chi^2(t,i,s,d)\ dt\ di$$

In general, there is some degeneracy in the (s,d) space, because closer, smaller objects will produce similar patterns to farther, larger objects. The diffraction effects are different, but the extent to which we can see the difference depends on the sampling rate and the signal to noise ratio.

Chapter 9

Multi Point Simulation

In this Chapter we consider test statistics that take into account more than one time point in the context of the normal model and the level $\alpha = 10^{-10}$. We illustrate a vulnerability of the *norm* statistic, then compare various detection methods for TAOS I and TAOS II. We find that despite their simplicity and limitations, the single point flux reduction for TAOS I and the norm statistic for TAOS II have good power for much of the parameter space of interest.

We focus on observations at opposition, Magnitude 11 and 13 stars with spectral type F0V, distances 43AU and 300AU, and our interest is in objects of sizes $0.5km - 10km$. In Figure 9.1 we plot a sample of occultation profiles at opposition, for each magnitude. The plots show a 10Hz sampling rate, and generate uniformly over size, distance, and impact parameter. We do not think the prior is uniform over size and distance, but we would like to have one test for a given star, and these plots show that we need not consider a range larger than $(-0.5, 0.5)$ seconds. Hence, we consider signals of lengths $5, 9, 19, 39, 79$ for sampling rates 5Hz, 10Hz, 20Hz, 40Hz, and 80Hz, respectively.

For some experiments below we use the representative signal to noise ratios

$$(M11, \sqrt{3} * 20) \ , \ (M13, \sqrt{3} * 7)$$

The term $\sqrt{3}$ comes from co-adding the three telescopes and the values 20 and 7 are motivated [1] by Figure 6.5.

9.1 A vulnerability of the *norm* statistic.

Here we imagine we have resolutions [2] ranging from $5Hz$ to $80Hz$ and their signal to noise ratios are related as if the data resulted from binning the $80Hz$ data. Since an M11 star at 5Hz is assumed to have signal to noise $\sqrt{3} * 20$, at $20Hz$ it will have signal to noise $4^{-.5} * \sqrt{3} * 20 = \sqrt{3} * 10$.

[1] The average in Figure 6.5 for M13 stars is 5, but that average includes many stars that have very little power, so we choose the slightly higher value of 7.

[2] We assume no timing parameter (defined in Chapter 1). Therefore, $10Hz$ resolution means the exposure time is 0.1 seconds.

In Figure 9.2 we look at the power of the norm statistic for these two magnitudes, and for distances $43AU$ and $300AU$. In most cases, higher resolution means lower power, which is an undesirable property. In a higher signal to noise case (not shown here), power sometimes increases when moving from 5Hz to 10Hz to 20Hz, but then eventually decreases for higher resolutions. In the case of very high signal to noise, the power is approximately 1 for all the resolutions.

9.2 Single Point VS Multi Point for TAOS

We investigate when it is worthwhile to consider procedures more complex than a single point flux reduction when observing at opposition. If we knew the prior distribution over size and distance, then for each magnitude star and each signal to noise ratio we could form the optimal test statistic, as described in Chapter 8.

We do not know the prior, so given the two representatives $(M11, \sqrt{3}*20)$, $(M13, \sqrt{3}*7)$, for each size and distance we create the optimal test statistic and calculate its power using importance sampling, as explained in section 8.2.3, giving the optimal power at each size and distance. The optimal power, along with the power of the single point flux reduction appears in Figure 9.3. We note that

- For each row in Figure 9.3, the optimal power given by the red curve is an upper bound bound for the optimal power of any level α test statistic designed for all sizes and distances, because each point on the red curve corresponds to a different optimal test statistic.
- The largest possible power increase over the single point test that **any** level α test can achieve is bounded by the biggest difference between the point-wise optimal test and the single point test. For the cases considered in Figure 9.3, the single point test does quite well.
- The experiment was done using different signal to noise ratios, and in some cases there is some potential to improve over the single point test. For very low signal to noise ratios, both tests have basically no power for small objects $< 4km$, but there appears to be some possible gain over the single point flux reduction for objects in the range $5km - 10km$. This is somewhat intuitive when looking at the 10Hz signals in Figure 9.1; when a signal has an 70% flux reduction for more than one point, the detection will be more significant when taking extra points into account, but the midpoint alone will be significant enough to pass the threshold as long as the signal to noise ratio is not too low.

9.3 Future Surveys: TAOS II

A second generation of the TAOS project (TAOS II) is being planned. TAOS II will have three telescopes operating at $20Hz$. Since CCD technology has advanced in the past 10 years, TAOS II will not run in Zipper mode and the signal to noise ratios will be much better.

For a star of given magnitude, how does the signal to noise ratio in 5Hz TAOS I data compare to the signal to noise ratio of the 20Hz TAOS II data? In general, how can we understand the relationship like that appearing in Figure 6.5?

If we approximate all non-Poisson noise as an additive background term with mean zero and variance b^2, independent of the Poisson noise, then our signal to noise ratio is given by

$$\lambda(\mu) = \frac{\mu}{\sqrt{\mu + b^2}}$$

The relationship between magnitude and the Poisson noise level is $\mu(m) = \mu(0) * 10^{-.4*m}$ where μ_0, the photons per unit time for a zero magnitude star, is a property of the photometric system and can be determined by calibration. This gives the approximation:

$$\lambda(m) = \frac{\mu(0) * 10^{-.4*m}}{\sqrt{\mu(0) * 10^{-.4*m} + b^2}}$$

which implies

$$.398 = 10^{-.4} \quad \leq \quad \frac{\lambda(m+1)}{\lambda(m)} \quad \leq \quad 10^{-.2} = 0.631$$

The upper bound is approached for bright stars ($b^2 << \mu$) while the lower bound is approached for dim stars ($b^2 >> \mu$). This is consistent with figure 6.5, where each increase in magnitude cuts the average signal to noise ratio approximately in half.

For TAOS II, the signal will overall increase by a factor of 3.2 through an increased aperture and different sampling rate. The background noise will dramatically decrease from the elimination of zipper mode (factor of ≈ 100). We assume [3] that the signal to noise ratio for a M13 star will increase by a factor of 40. Using the TAOS I averages of 5 and 20 for M13 and M11, and assuming $\frac{\lambda(m+1)}{\lambda(m)} = \frac{1}{2}$ we get parameters

$$(M11, \sqrt{3} * 800) \ , \ (M13, \sqrt{3} * 200)$$

We consider several 19 point detection methods for this 20Hz data. As in Figure 9.3, we consider the a test based on the Neyman-Pearson lemma for each size and distance, and the single point flux reduction at the center of the 19 point. We also consider windows around the central point with sizes $3, 5, 7, ..., 19$, and consider the average and the norm inside that window.

In Figure 9.4, we look at the power of these methods. **BestMA** is the most powerful of all the moving average tests over the possible windows, and **BestChi** is the best of all the tests based on the norm test statistic. **ChiFull** is the test based on the norm of all 19 points. We note, that **SP** and **ChiFull** are the only tests listed that use the same test statistic for all tests considered. We note that

- Unlike the cases in Figure 9.3, there is a lot to gain over the single point flux reduction.

[3] All ballpark estimates here have been based on personal communications with Matt Lehner. We then err on the safe side; values chosen here are on the small side of those ballpark estimates.

- Optimizing over moving averages does not give much advantage over the single point test.
- The statistic **ChiFull** is a big improvement over the moving average.
- At 43AU, **ChiFull** has essentially perfect power for all cases considered. Therefore, there is no advantage to considering complex methods if searching these stars for objects at 43 AU.
- At 300AU, **ChiFull** performs almost the same as the more complicated **BestChi**, and it has high power for objects of size $> 2km$.
- At 300AU, there is some room for improvement over **ChiFull** for small objects.

It may seem odd that the power curves are not monotone in size. However, sometimes diffraction effects dominate the signal, so there are instances where slight increases in the diameter of the occulting object make the diffraction profile less pronounced.

9.3.1 Dim Stars

It is known that the luminosity function [5] is approximated by a power law with slope 2.5, so increasing the limit by 1 (eg M14 to M15) will result in 2.5 times as many stars [5]. The plan for TAOS II is to have a lower limiting magnitude of 17.5 (rather than 13.5 in TAOS), and since $2.5^4 \approx 40$, without crowding about $\frac{40}{41}$ of the target stars would have magnitudes in $(13.5, 17.5)$. For this reason, we repeat the above experiment for dimmer stars, again using the relationship $\frac{\lambda(m+1)}{\lambda(m)} = \frac{1}{2}$ to get parameters

$$(M15, \sqrt{3} * 50) \quad , \quad (M17, \sqrt{3} * 12.5)$$

We also consider another method, **BestPCA**, based on the theory in Section 8.3.3. For each magnitude, size, and distance,

1. Generate M impact parameters uniformly and create the $M \times 19$ matrix of templates.
2. For each $k \in 1, ..., 19$ let X be the first k principal components and calculate the power based on the linear regression test described in Section 8.3.3
3. Keep the power corresponding to the best choice of k.

Figure 9.5 shows the power curves for the methods and Figure 9.6 is a truncated version which only shows results for small objects. We note

- For M15 stars and objects $\geq 1.5km$, **ChiFull** has nearly perfect power.
- For small and distant objects, there may be some improvement over **ChiFull**.
- For M17 stars, the method **BestChi** performs a little better than **ChiFull**.

[5] This is the cumulative distribution of the number of stars.

[5] Due to crowding in the field, we cannot observe all of them so the effective increase will be less.

- The method **BestPCA** performs almost identically to the likelihood ratio test.
- For small objects occulting M17 stars, there appears to be some room for improvement over **ChiFull**, and the method **BestPCA** shows some potential for improvement.

9.3.2 Practical Issues and Considerations

Above we acted as if we averaged three independent telescopes to get a $\sqrt{3}$ improvement in signal to noise. In practice it may be more desirable and robust to combine three independent tests, such as by using the rank product test statistic. For example, we could take a 19-point running norm of each of the time series, rank them from highest to lowest, then look at the rank product.

Outliers may be a bigger problem when using something like the running norm because an outlier at one point could turn into outliers at 20 points. Depending on the relationship between the lengths of the series and outliers, this could hurt the power of the rank product method.

Although computationally expensive, the $\Delta\chi^2$ statistic based on a complicated objective function could be used. In particular, the diffraction profile is a function of three parameters: size, distance, and impact parameter. It might also be worth considering robust optimization; $\Delta\chi^2$ uses least squares, but one could use trimmed least squares or the $L1$ norm. Significance could then be assessed by permutation methods, as was done in [Schlichting et al., 2009]. An alternative to permutation tests would be to take a running $\Delta\chi^2$ on each telescope, then rank the data and look at the rank product. It is possible that the optimization would choose different parameters on each telescope giving a significant rank product, but that does not invalidate the significance result.

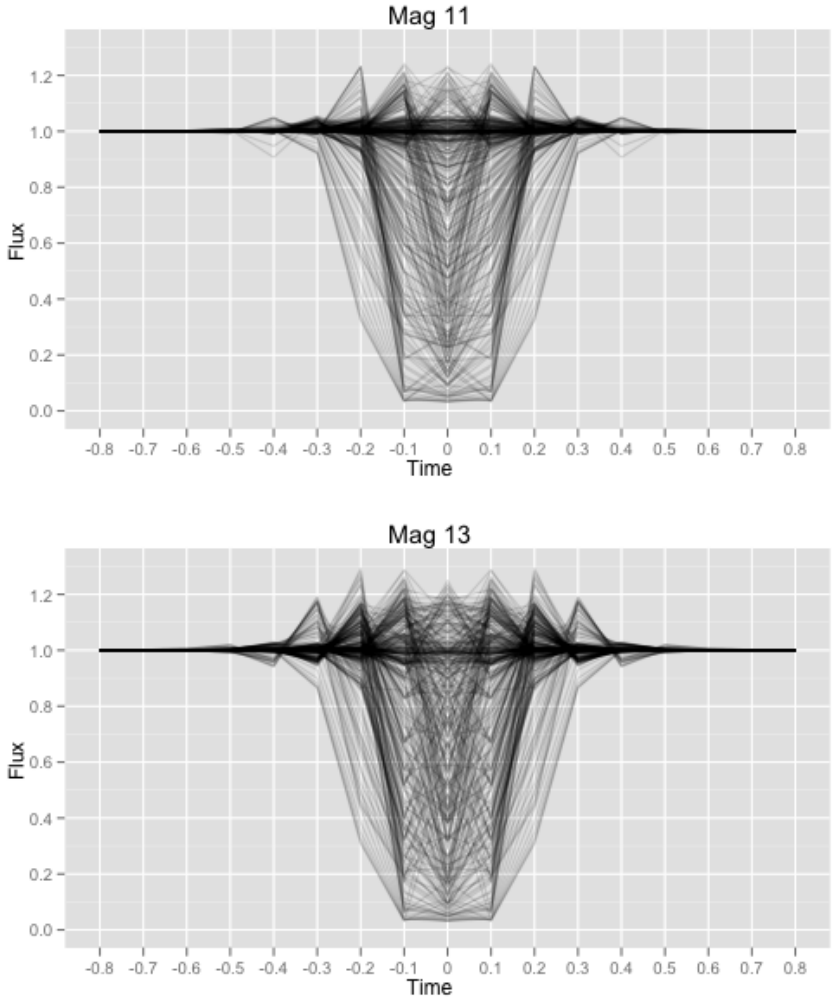

Figure 9.1: 10Hz occultation profiles at opposition. Profiles are a uniform random sample over distances 43AU and 300AU, sizes .5 km - 10 km, and the impact parameter.

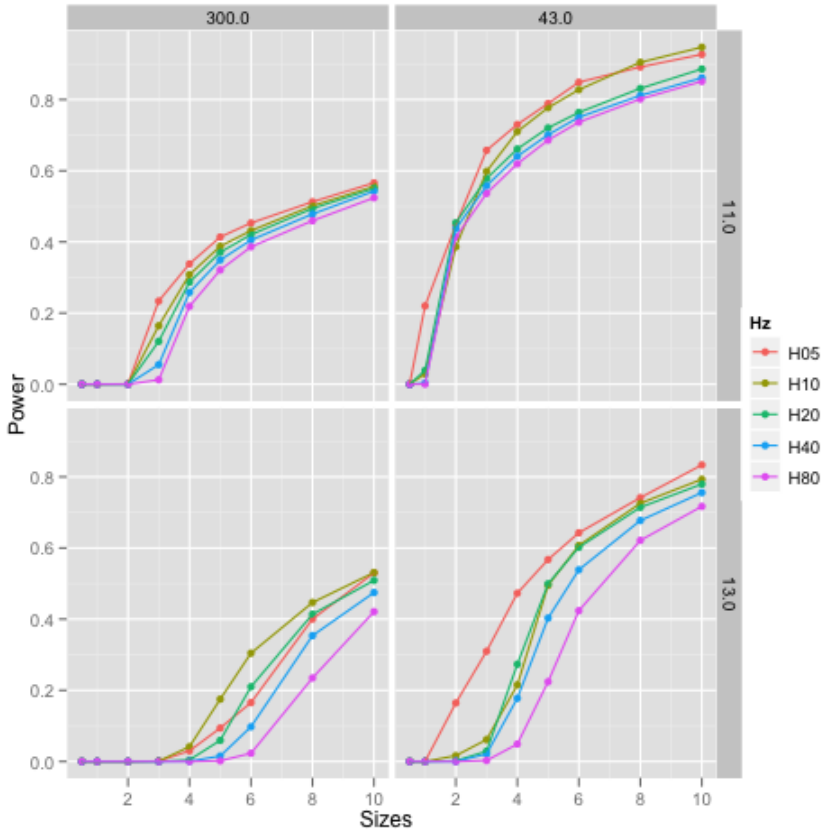

Figure 9.2: The Power of detection using the norm statistic. Signal to noise rates are adjusted as if they arose from binning independent data. Rows correspond to magnitudes while columns correspond to distances.

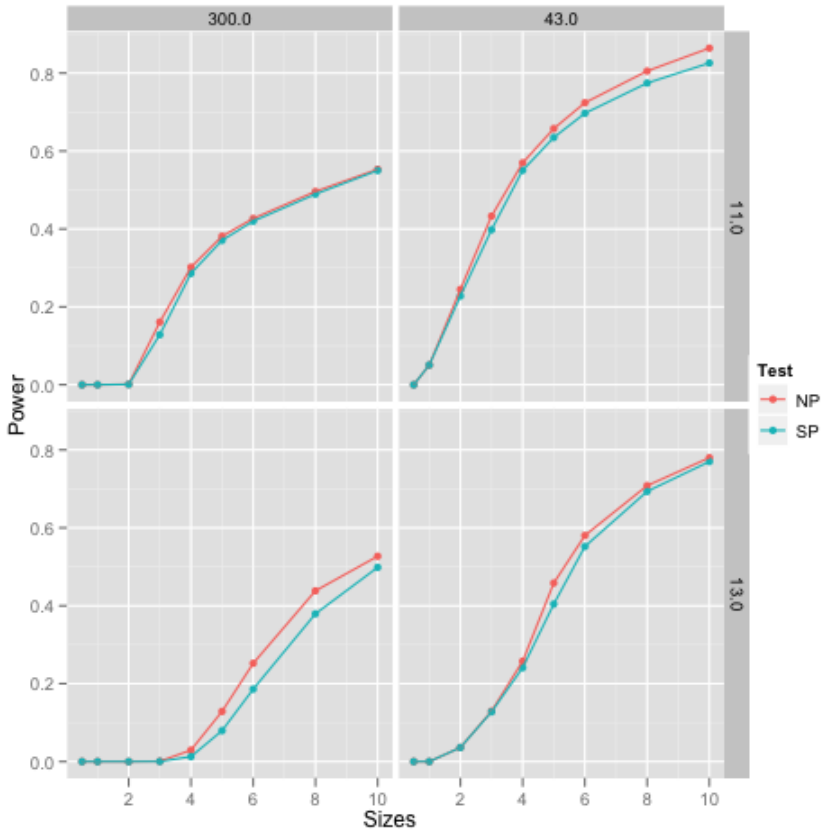

Figure 9.3: Power of the single point flux reduction and of the Neyman-Pearson lemma based on five points in the 5Hz data. Signal to noise ratios of $\sqrt{3} * 7$ and $\sqrt{3} * 20$ are assumed for $M13$ and $M11$ stars, respectively. Rows correspond to magnitudes while columns correspond to distances.

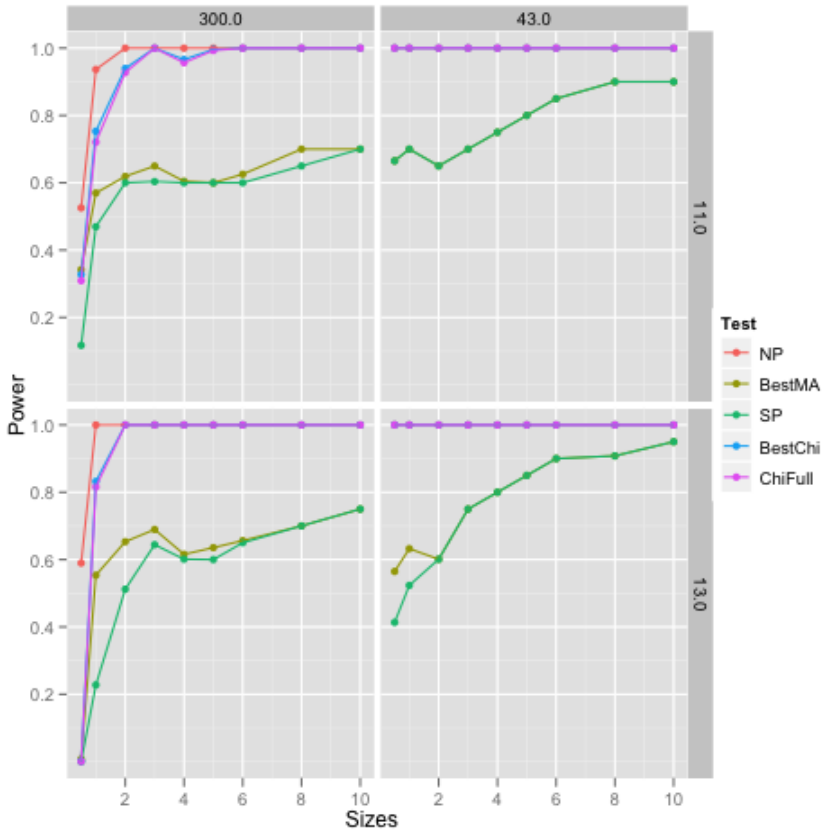

Figure 9.4: Power of methods based on 19 points in the 20Hz data. Signal to noise ratios of $\sqrt{3}*200$ and $\sqrt{3}*800$ are assumed for $M13$ and $M11$ stars, respectively. Rows correspond to magnitudes while columns correspond to distances. There is over-plotting here, but curves are plotted in the order listed in the left column. Since we must have NP > BestChi > ChiFull and they are plotted in that order, when one color is missing we know it is under the later colors. It is also the case that BestMA is often under SP.

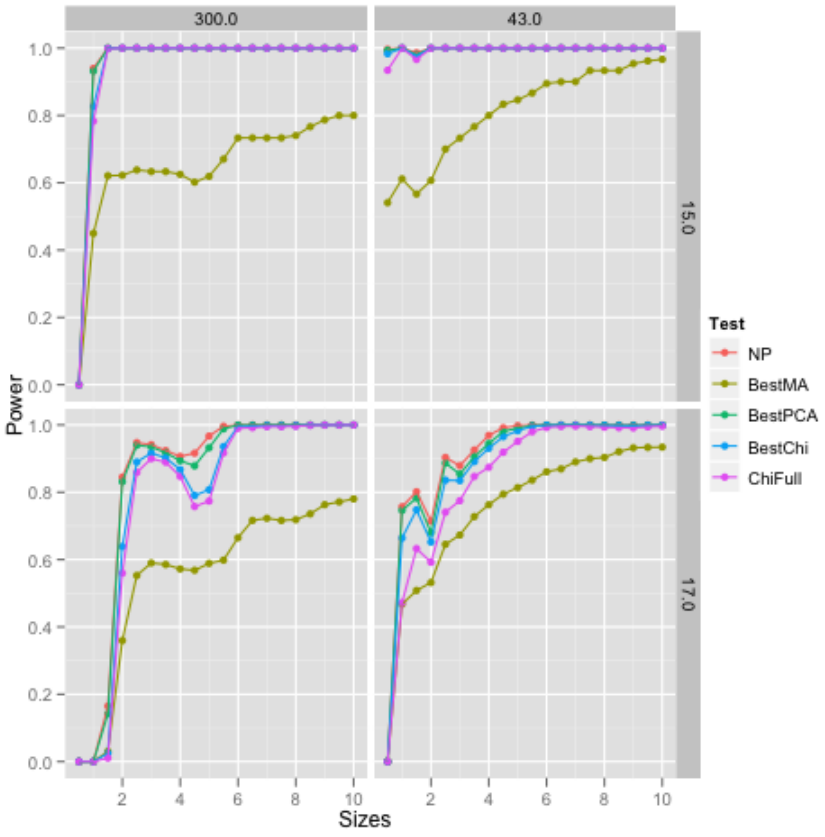

Figure 9.5: Comparison of methods for the dim stars with approximate TAOS II parameters.

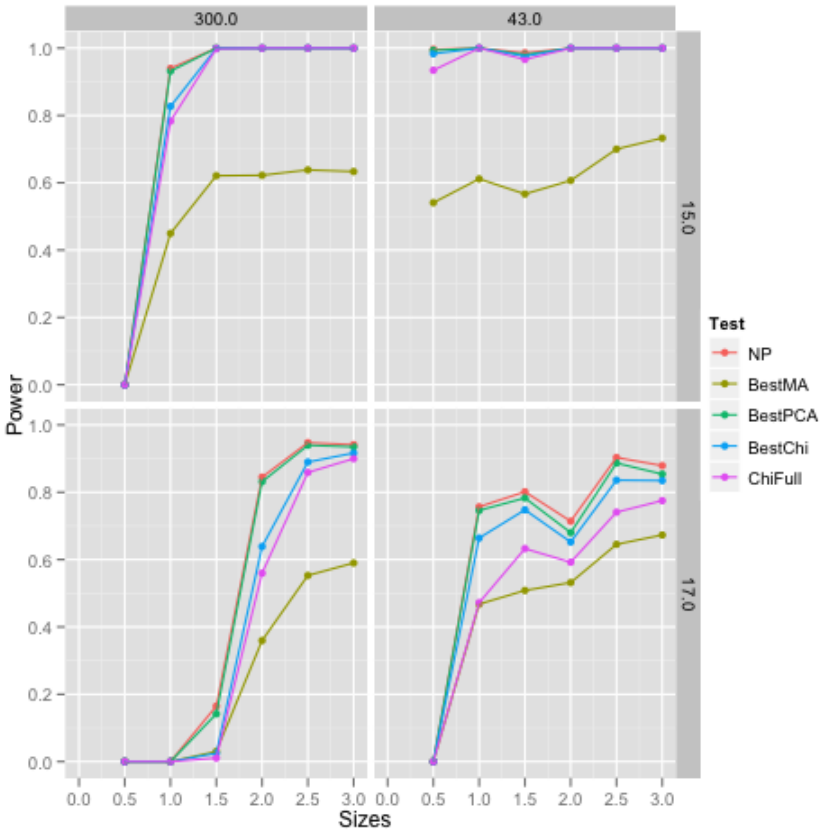

Figure 9.6: View of Figure 9.5 for small sizes.

Chapter 10

Appendix

10.1 Union Vs Exact

Here we examine the relative error comparing the union bound of rank products with the minimum rank product. We assume the data is exchangeable so all permutations are equally likely, and continuous so there are no ties.

10.1.1 Distribution of the Minimum

Let r_i^k be the rank at time i on the k'th telescope. Let the rank product be $R_i = \prod_{k=1}^{K} r_i^k$. Let X be the statistic of interest: $X = \min_i R_i$. Then

$$\begin{aligned}
\mathbb{P}(X \leq x) &= \mathbb{P}(\cup_{i=1}^{N}\{R_i \leq x\}) \\
&= \sum_{j=1}^{N}(-1)^{j+1} \sum_{I\subset 1,\ldots,n \quad |I|=j} \mathbb{P}(\cap_{i\in I}\{R_i \leq x\}) \\
&= \sum_{j=1}^{N}\binom{N}{j}(-1)^{j+1}\mathbb{P}(\cap_{i=1}^{j}\{R_i \leq x\}) \\
&= \sum_{j=1}^{x}\binom{N}{j}(-1)^{j+1}\mathbb{P}(\cap_{i=1}^{j}\{R_i \leq x\}) \\
&= N * \mathbb{P}(R_1 \leq x) + \sum_{j=2}^{x}(-1)^{j+1}\binom{N}{j}\mathbb{P}(\cap_{i=1}^{j}\{R_i \leq x\})
\end{aligned}$$

The second equality is the inclusion exclusion principle and the third follows from exchangeability. The fourth equality follows since higher order intersections are empty. After all, the first coordinate of the rank tuple must be different at each time point and must take values in $1, 2, \ldots x$.

10.1.2 Bounding the Relative Error

We are interested in the relative error of the union bound in computing the distribution of the minimum:

$$RE = \frac{\text{Approximation} - \text{Truth}}{\text{Truth}} = \frac{N * \mathbb{P}(R_1 \leq x) - \mathbb{P}(X \leq x)}{\mathbb{P}(X \leq x)}$$

So if L is a lower bound for $\mathbb{P}(X \leq x)$, then we have

$$0 \quad \leq \quad RE \quad \leq \quad \frac{N * \mathbb{P}(R_1 \leq x)}{L} - 1$$

We note that if $J \leq x$ is even, then a lower bound L is given by

$$N * \mathbb{P}(R_1 \leq x) + \sum_{j=2}^{J} (-1)^{j+1} \binom{N}{j} \mathbb{P}(\cap_{i=1}^{j} \{R_i \leq x\})$$

It turns out for this problem that $J = 2$ gives a good enough bound.

10.1.3 Computation of Intersections

We assume now $x \leq N$, that is, we only consider rank products smaller than the length of the series in question. This includes the probabilities of interest and allows us to focus on counting problems that will be valid for series of many different lengths. The intersections can be written

$$\begin{aligned} \mathbb{P}(\cap_{i=1}^{j} \{R_i \leq x\}) &= \frac{1}{N^K (N-1)^K ... (N-(j-1))^K} * f(x, K, j) \\ &= (\frac{(N-j)!}{N!})^K * f(x, K, j) \end{aligned}$$

Where $f(x, K, j)$ is the number of ways to choose K ordered samples of size j from 1-N (or 1-x), take the pointwise product then take the minimum and have it be less than x.

We note that for any such sample, we could permute the coordinates and get another, so we define $g(x, K, j) = \frac{f(x,K,j)}{j!}$

We combine this with the above and it becomes a counting problem:

$$\begin{aligned}
\mathbb{P}(X \leq x) &= N * \mathbb{P}(R_1 \leq x) + \sum_{j=2}^{x} (-1)^{j+1} \binom{N}{j} \mathbb{P}(\cap_{i=1}^{j} \{R_i \leq x\}) \\
&= (\frac{1}{N})^{K-1} f(x, K, 1) + \sum_{j=2}^{x} (-1)^{j+1} \binom{N}{j} (\frac{(N-j)!}{N!})^{K} * f(x, K, j) \\
&= (\frac{1}{N})^{K-1} g(x, K, 1) + \sum_{j=2}^{x} (-1)^{j+1} \frac{N!}{j!(N-j)!} (\frac{(N-j)!}{N!})^{K} * j! * g(x, K, j) \\
&= (\frac{1}{N})^{K-1} g(x, K, 1) + \sum_{j=2}^{x} (-1)^{j+1} (\frac{(N-j)!}{N!})^{K-1} * g(x, K, j)
\end{aligned}$$

10.1.4 Bounding Relative Error by g

Using the second order term, we bound the relative error by:

$$\begin{aligned}
RE &= \frac{(\frac{1}{N})^{K-1} g(x, K, 1) - \mathbb{P}(X \leq x)}{\mathbb{P}(X \leq x)} \\
&= \frac{(\frac{1}{N})^{K-1} g(x, K, 1)}{\mathbb{P}(X \leq x)} - 1 \\
&\leq \frac{(\frac{1}{N})^{K-1} g(x, K, 1)}{(\frac{1}{N})^{K-1} \left(g(x, K, 1) - (\frac{1}{(N-1)})^{K-1} * g(x, K, 2) \right)} - 1 \\
&= \frac{(\frac{1}{N-1})^{K-1} g(x, K, 2)}{g(x, K, 1) - (\frac{1}{N-1})^{K-1} g(x, K, 2)}
\end{aligned}$$

10.1.5 Computation of g

The functions $g(x, K, j)$ are evaluated with a simple R script. The method would be hard for large x, but for tail behavior of interest, $x < 150$ is sufficient and the computations to make figure 2.2 takes less than 70 seconds on a laptop.

A simple example

We use a for loop to collect all tuples corresponding to a rank product less than or equal to x. For example, if we have three telescopes and $x = 4$ we have

1	1	1
1	1	2
1	1	3
1	1	4
1	2	1
1	2	2
1	3	1
1	4	1
2	1	1
2	1	2
2	2	1
3	1	1
4	1	1

This list shows that $g(4,3,1) = 13$. To find $g(4,3,2)$ we would count all compatible pairs, where compatible means the triples could appear in the same time series.

We can think of a 'Compatibility Graph' on these 13 points: a is connected to b iff a and b are compatible iff $a[1] \neq b[1], a[2] \neq b[2], a[3] \neq b[3]$,

In such a graph, if we had a clique of size $M > 2$, then those M points could appear in the same time series, and the number of cliques of size M would give us $g(4,3,M)$. The undirected graph can be viewed as a 13 by 13 upper-triangular matrix:

$$\begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}$$

- Hence $g(4,3,2) = 9 =$ sum of elements of the matrix above
- $g(4,3,3) = g(4,3,4) = 0$ because there are no cliques of size 3 or higher

$$\begin{aligned}
\mathbb{P}(X \leq x) &= (\frac{1}{N})^{3-1}\left(13 - (\frac{1}{(N-1)})^{3-1} * 9 + 0\right) \\
&= \frac{1}{N^2}(13 - 9 * \frac{1}{(N-1)^2})
\end{aligned}$$

If N=4, this is $1/16 * (13 - 9/9) = .75$

If N=7, this is $1/49 * (13 - 9/36) = .26$

10.2 The $LogSumExp()$ function

Suppose v is a vector of length M. Then for any $1 \leq j \leq M$

$$\begin{aligned} LogSumExp(v) &\equiv log(\sum_{i=1}^{M} \exp(v_i)) \\ &= v_j + log(\sum_{i=1}^{M} \exp(v_i - v_j)) \\ &= v_j + LogSumExp(v - v_j) \end{aligned}$$

Where $v - v_j$ in the last line means subtract the constant v_j from each element of the vector v.

A problem with naive computation of $LogSumExp$ is that things can blow up when taking exponentials. However, if we first find the max ($v^* = max_j\{v_j\}$) and use the equation above we get rid of this problem because every element of $v - v^*$ is non-positive.

This gives the upper bound

$$\begin{aligned} LogSumExp(v) &= v^* + LogSumExp(v - v^*) \\ &\leq v^* + LogSumExp(v - v^*) \\ &= v^* + log(M) \end{aligned}$$

This becomes an equality if all the elements are equal.

For the lower bound, we note that since the j'th element of $v - v_j$ is zero, we must have $SumExp(v - v_j) > 1$, so $LogSumExp(v - v_j) > 0$ which implies that

$$v^* < LogSumExp(v)$$

This is never exact for $M > 1$. We can get a tighter bound by assuming that the k'th largest element is at distance greater than $d > 0$ from the max

$$v^* < LogSumExp(v) \leq v^* + log\Big((k-1) + (M-k+1)\exp(-d)\Big)$$

So it gets arbitrarily close to the max if the max is well separated from the rest: let $k = 2$ and $d \to \infty$.

10.3 Importance Sampling the tails of LLR

10.3.1 Importance Sampling: The basic idea

In general, suppose we X has density f and we are interested in

$$\theta = \mathbb{E}(h(x)) = \int h(x)f(x)dx$$

Suppose we generate Y according to density g, then we have

$$\begin{aligned}\mathbb{E}_x(h(x)) &= \int h(x)f(x)dx \\ &= \int \frac{h(y)f(y)}{g(y)}g(y)dy \\ &= \mathbb{E}_y(\frac{h(y)f(y)}{g(y)})\end{aligned}$$

Where we are assuming the support of g contains that of f.

This suggests that one could generate $Y_1, ..., Y_n$ independently with density g, and then get an unbiased estimate of θ by

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}\frac{h(Y_i)f(Y_i)}{g(Y_i)}$$

10.3.2 Tail Probabilities for the LLR

Suppose we observe a random vector X and our test statistic is the log-likelihood ratio, $LLR(X)$. For some c we wish to estimate

$$\theta = \mathbb{P}(LLR(X) \geq c) = \mathbb{E}(h(X)) \qquad \text{Where} \qquad h(x) = \mathbf{1}_{LLR(X)\geq c}$$

Suppose we have made m independent copies of the $LLR(Z_i)$ where we generated the vector Z_i according to the alternative. Suppose we have collected the result in a vector A of length m and that the results are ordered from largest to smallest. Then

$$\begin{aligned}\hat{\theta}(c) &= \frac{1}{m}\sum_{i=1}^{m}\frac{1}{LR(Z_i)}\mathbf{1}_{LLR(Z_i)>c} \\ &= \frac{1}{m}\sum_{i=1}^{m}\exp(-LLR(Z_i))\mathbf{1}_{LLR(Z_i)>c} \\ &= \frac{1}{m}\sum_{i=1}^{k(c)}\exp(-A[i]) \qquad \text{Where} \qquad k(c) = max\{i : A[i] \geq c\}\end{aligned}$$

10.3.3 Simulation for the Normal model

Consider the Model

$$z = C(\theta) + \frac{1}{\lambda}\epsilon \qquad for \quad \epsilon \sim \mathcal{N}(0, I_n)$$

And assume θ has a prior over $\{\theta_1, ..., \theta_M\}$ with probabilities $p_1, ..., p_M$. Let C denote the $M \times n$ matrix where rows correspond to signals and let $\Omega = CC'$.

Using the derivation from chapter 8, we see that under alternative j we have

$$\begin{aligned} LLR_j(z) &= LogSumExp\left(log(p_i) + \frac{-\lambda^2}{2} C(\theta_i)'C(\theta_i) + \lambda^2 < z, C(\theta_i) > \right) \\ &= LogSumExp\left(log(p_i) + \frac{-\lambda^2}{2} \Omega[i,i] + \lambda^2 \Omega[j,i] + \lambda < \epsilon, C(\theta_i) > \right) \\ &\equiv LLR(j, \epsilon, C) \end{aligned}$$

Hence, we can generate an iid sample of $LLR(z)$ under the alternative by iterating the procedure

1. Generate $j \in \{1, ..., M\}$ according to probabilities $p_1, ..., p_M$.
2. Generate $\epsilon \sim \mathcal{N}(0, I_n)$ independently of j.
3. Calculate $LLR(j, \epsilon, C)$

Bibliography

[Bernstein et al., 2004] Bernstein, G., Trilling, D., Allen, R., and Brown, M. (2004). The size distribution of trans-Neptunian bodies. *The Astronomical Journal*.

[Bianco, 2009] Bianco, F. (2009). Chasing Shadows in the Outer Solar System. *PhD Thesis*.

[Bianco et al., 2010] Bianco, F., Zhang, Z., and Lehner, M. (2010). The TAOS Project: Upper Bounds on the Population of Small Kuiper Belt Objects and Tests of Models of Formation and Evolution of the Outer Solar System. *The Astronomical Journal*.

[Box et al., 2005] Box, G., Hunter, J., and Hunter, W. (2005). *Statistics for Experimenters*.

[Durrett, 2005] Durrett, R. (2005). *Probability: Theory and Examples*. Duxbury Press, third edition.

[Fisher, 1956] Fisher, R. A. (1956). *Statistical Methods for Research Workers*. Oliver and Boyd.

[Freedman, 2005] Freedman, D. (2005). *Statistical Models: Theory and Practice*. Cambridge University Press, first edition.

[Ibragimov and Linnik, 1971] Ibragimov, I. and Linnik, Y. (1971). *Independent and Stationary Sequences of Random Variables*.

[Kim et al., 2009] Kim, D.-W., Protopapas, P., Alcock, C., Byun, Y.-I., and Bianco, F. B. (2009). Detrending time series for astronomical variability surveys. *Monthly Notices of the Royal Astronomical Society*, 397(1):558–568.

[Kovacs et al., 2005] Kovacs, G., Bakos, G., and Noyes, R. (2005). A trend filtering algorithm for wide-field variability surveys. *Monthly Notices of the Royal Astronomical Society*.

[Kunsch, 1989] Kunsch, H. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, pages 1217–1241.

[Lehner et al., 2010] Lehner, M., Coehlo, N., Zhang, Z., Bianco, F., Wang, J., Rice, J., Protopapas, P., Alcock, C., Axelrod, T., and Byun, Y. (2010). The TAOS Project: Statistical Analysis of Multi-Telescope Time Series Data. *Arxiv preprint arXiv:1002.3626*.

[Lehner et al., 2009] Lehner, M., Wen, C., and Wang, J. (2009). The Taiwanese-American Occultation Survey: the multi-telescope robotic observatory. *Publications of the Astronomical Society of the Pacific*.

[Liang, 2001] Liang, C. (2001). The detection of stellar occultations by Kuiper Belt objects. *PhD Thesis*.

[Liang et al., 2004] Liang, C., Rice, J., de Pater, I., Alcock, C., Axelrod, T., Wang, A., and Marshall, S. (2004). Statistical Methods for Detecting Stellar Occultations by Kuiper Belt Objects: the Taiwanese-American Occultation Survey. *Statist. Sci*.

[Mosteller and Fisher, 1948] Mosteller, F. and Fisher, R. A. (1948). Questions and Answers. *The American Statistician*, 2(5):30–31.

[Nihei et al., 2007] Nihei, T., Lehner, M., Bianco, F., and King, S. (2007). Detectability of Occultation of Stars by Objects in the Kuiper Belt and Oort Cloud. *The Astronomical Journal*.

[Rudin, 1976] Rudin, W. (1976). *Principles of Mathematical Analysis*. third edition.

[Schlichting et al., 2009] Schlichting, H. E., Ofek, E. O., Wenz, M., Sari, R., Gal-Yam, A., Livio, M., Nelan, E., and Zucker, S. (2009). A single sub-kilometre Kuiper belt object from a stellar occultation in archival data. *Nature*, 462(7275):895–897.

[Sen, 1989] Sen, P. K. (1989). The Mean-Median-Mode Inequality and Noncentral Chi Square Distributions. *Sankhya: The Indian Journal of Statistics, Series A*, 51(1):106–114.

[Wang et al., 2009] Wang, J., Lehner, M., Zhang, Z., and Bianco, F. (2009). Upper Limits on the Number of Small Bodies in Sedna-Like Orbits. *The Astronomical Journal*.

[Zhang, 2009] Zhang, Z. (2009). Automated Photometry Pipeline and Event Detection Algorithm in the TAOS Project. *thesis.lib.ncu.edu.tw*.

[Zhang et al., 2008] Zhang, Z., Bianco, F., Lehner, M., and Coehlo, N. (2008). First Results from TAOS. *The Astronomical Journal*.

[Zhang et al., 2009] Zhang, Z., Kim, D., Wang, J., and Lehner, M. (2009). The TAOS Project: High-Speed Crowded Field Aperture Photometry. *Publications of the Astronomical Society of the Pacific*.