

TESTING ONE HYPOTHESIS MULTIPLE TIMES

A THESIS PRESENTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY OF IMPERIAL COLLEGE LONDON
AND THE
DIPLOMA OF IMPERIAL COLLEGE
BY
SARA ALGERI

DEPARTMENT OF MATHEMATICS
IMPERIAL COLLEGE
180 QUEEN'S GATE, LONDON SW7 2AZ

APRIL 2018

I certify that this thesis is the product of my own work and that any ideas or quotations from the work of other people, published or otherwise, are properly acknowledged.

Sara Algeri

COPYRIGHT

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Thesis advisor: Prof. David A. van Dyk
Thesis co-advisor: Prof. Jan Conrad

Sara Algeri

Testing One Hypothesis Multiple Times

ABSTRACT

The identification of new rare signals in data, the detection of a sudden change in a trend, and the selection between competing models, are among the most challenging problems in statistical practice. In these settings, standard regularity conditions (e.g., those required by Wilks theorem) fail to hold and thus classical inferential tools, such as the generalized likelihood ratio test, are not applicable. In this thesis, we show how these challenges can be tackled using a test of hypothesis where a nuisance parameter is present only under the alternative. Several solutions have been proposed in the statistical literature and their practical implementation often reduces the problem into one of Testing One Hypothesis Multiple times (TOHM). Specifically, a fine discretization of the space of the non-identifiable parameter is specified, and an ensemble of sub-test statistics is obtained to test the null hypothesis against a set of sub-alternative hypothesis where the value of the non-identifiable parameter is fixed at each point of the discretization. The goal is to provide a global p-value as the standard of evidence for comparing the null hypothesis and the alternative hypothesis. In this thesis, we combine elements of extreme value theory, differential geometry, graph theory and simulations methods to provide an easy to compute, computationally efficient and highly generalizable inferential tool to perform TOHM under stringent significance requirements, such as those typically required in the physical sciences. The methods proposed in this thesis formalize and extend recent results presented in physics literature. Several applications are discussed in the context of indirect searches for dark matter.

LIST OF PUBLICATIONS

This thesis collects results and methods originally published in the following manuscripts.

Algeri S. and van Dyk D. A. (2018). Testing one hypothesis multiple times: the multidimensional case. *Submitted. arXiv:1803.03858.*

Algeri S. and van Dyk D. A. (2017). Testing one hypothesis multiple times. *Submitted. arXiv:1701.06820.*

Algeri S., van Dyk D. A., Conrad J., Anderson B. (2016). On methods for correcting for the look-elsewhere effect in searches for new physics. *Journal of Instrumentation 11 P12010.*

Algeri S., Conrad J., van Dyk D. A. (2016). A method for comparing non-nested models with application to astrophysical searches for new physics. *Monthly Notices of the Royal Astronomical Society: Letters 458 (1), L84-L88.*

ACKNOWLEDGMENTS

I owe a great debt of gratitude to many people for the help and support I was blessed with during the writing of this thesis.

With nostalgia for the upcoming end of this era, I thank my advisor David for believing in me through all the challenges I have undertaken in these four years; if me or my work were able to flourish on any of these occasions, it was in major part thanks his guidance. I am also extremely grateful to my co-advisor Jan, whose encouragement has been crucial to acquire strength and confidence from the very beginning. The discussions with both my advisors have been an essential source of motivation for my work and strongly contributed to my growth as a scientist and a scholar.

Among the many friends and colleagues I met during my studies at Imperial College, I wholeheartedly thank Andrea, co-protagonist of an unforgettable PhD experience and great friend. I say thank you to Din, who was always there to assist me, and to Heather, for her confidence in me and her support when I most needed it. I also thank my colleagues in Office 6M09 for the simple things we enjoyed together, and the Aeronautics group for the exhilarating conversations governed by freedom of speech at its best.

The preparation of this thesis would have not been possible without the understanding and support of my family during my most stressful and busy periods, and in particular my mom, who always did her best to make my trips back home special, and who I acknowledge with profound gratitude.

Finally, I say thank you to Deep, my invisible influencer and my main source of inspiration. Thanks to him I have learned how not to be afraid to take risks and the importance of setting high standards. He has never failed to prove to me that the gap between impossible and improbable can be filled by hard work and good taste, and that good ideas are the most valuable currency. I do not know what the future holds for me, but for better or for worse, I will always cherish his precious lessons.

LIST OF FIGURES

1.1	Upcrossings of $\{T(\theta)\}$	10
2.1	Types of dark matter searches	21
2.2	Model selection via LRT	26
2.3	Type I error and power for Example 2.1	27
3.1	Data and fitted models for Examples 2.1, 3.1 and 3.2	36
3.2	Simulated traces and upcrossings plot for Example 2.1	39
3.3	Simulated traces and upcrossings plots for Examples 3.1 and 3.2	40
3.4	Approximating excursion probabilities via TOHM for Examples 2.1, 3.1 and 3.2	42
4.1	Upcrossings and exceedances	45
4.2	Assessing Berman's condition	52
5.1	Null and sub-alternative models for Examples 3.1 and 5.1	56
5.2	Type I error and power for Examples 3.1a, 3.1b, 5.1	58
5.3	Implementing TOHM and MHT on Fermi LAT data	61
5.4	Guidelines for statistical signal detections via MHT and TOHM	63
5.5	QQ-plots for local and global p-values	66
6.1	Euler Characteristic in two dimensions	70
6.2	Quadrilateral mesh, hypercubes, graph and cliques	77
6.3	Data and fitted models for Examples 6.1, 6.2 and 6.3	81
6.4	Approximating excursion probabilities via TOHM for Examples 6.1, 6.2 and 6.3	82

LIST OF TABLES

1.1	Physics discoveries and σ -levels	11
2.1	Non-nested models comparison on Fermi LAT simulated data	28
4.1	Data analysis for Examples 2.1, 3.1 and 3.2	53
5.1	Implementing TOHM and MHT on Fermi LAT data	60
5.2	Type I error and power of the sequential approach	64
5.3	Computational time of the sequential approach	65
6.1	EC densities for Gaussian and χ^2 random fields	71
6.2	Data analysis for Examples 6.1, 6.2 and 6.3	84

CONTENTS

PREFACE	1
1 BACKGROUND ON HYPOTHESIS TESTING UNDER NON-STANDARD REGULARITY CONDITIONS	5
1.1 The large-samples distribution of the MLE and LRT	5
1.1.1 Classical regularity conditions	5
1.1.2 Testing under non-standard regularity conditions	6
1.2 Volume-of-tube formulae	13
1.3 Detecting signals via Multiple Hypothesis Testing	17
2 TESTING NON-NESTED MODELS VIA THE LIKELIHOOD RATIO TEST	20
2.1 Non-nested models comparison in dark matter searches	20
2.2 Statistical comparison of non-nested models	22
2.3 Validation on dark matter models and application to simulated data from the Fermi-LAT	25
3 TESTING ONE HYPOTHESIS MULTIPLE TIMES: THE ONE-DIMENSIONAL CASE	30
3.1 Definition and formalization	30
3.1.1 TOHM bounds for Gaussian-related processes	32
3.1.2 Testing One Hypothesis Multiple times in practice	35
3.2 Practical matters	37
3.2.1 Case studies: description	37
3.2.2 The choices of c_0 and R	38
4 TESTING ONE HYPOTHESIS MULTIPLE TIMES AND MULTIPLE HYPOTHESIS TESTING: TWO SIDES OF THE SAME COIN	44
4.1 Review on main results on point processes of upcrossings and exceedances	44

4.2	TOHM via Bonferroni's correction	47
4.2.1	Assessing D and D' for Gaussian and related sequences	49
4.3	TOHM, Bonferroni and Berman's condition	51
4.3.1	Data analyses	53
5	PRACTICAL GUIDELINES TO SELECT AMONG DIFFERENT INFERENCE PROCEDURES	55
5.1	Choice of the test statistics and statistical properties	56
5.2	TOHM and MHT on Fermi LAT data	61
5.3	A sequential approach to select among different procedures	62
6	TESTING ONE HYPOTHESIS MULTIPLE TIMES: THE MULTIDIMENSIONAL CASE	68
6.1	A quick review on random fields and geometry	68
6.2	Methodological setup	72
6.2.1	LRT and global p-values when testing on the boundary	75
6.3	Computing the mean Euler characteristic via graphs	75
6.4	Numerical results	80
6.4.1	Motivating Examples	80
6.4.2	Goodness of the approximations and choice of c_k	83
6.4.3	Data analysis	85
	CONCLUSION	86
	References	90
	APPENDIX A RATES OF CONVERGENCE AND GOODNESS OF THE APPROXIMATIONS	97
A.1	Rates of convergence	97
A.2	Approximating TOHM via Bonferroni	99
	References in Appendix	101

PREFACE

The data revolution characterizing the 21st century has led to a massive increase in the demand of modern analytical tools to effectively extrapolate useful and exhaustive information from the volume, the variety and the complexity of the data available. The need of statistical innovation has contributed to exponential advancements in the areas of data mining, machine learning and data science, and the vast variety and complexity of new scientific problems has encouraged the development of statistical solutions which generalize beyond the specifics of the original scientific settings. In this thesis, we attempt to address this demand by providing a generalized framework to perform statistical inference in the context scientific discoveries.

In statistical terms, discoveries of new phenomena often translate into a problem of identifying an unexpected mode in the distribution of data, detecting a sudden change in a data trend or selecting among competing non-nested models. The main difficulty of tackling this class of problems with classical inferential procedures is that standard asymptotic results [e.g., [Wilks, 1938](#)] do not apply. Whereas, solutions based on simulations and resampling methods such as the bootstrap [e.g., [Efron and Tibshirani, 1994](#)] may become prohibitive in a multidimensional framework, when dealing with complex models or under stringent significance requirements. This aspect is particularly relevant in high energy physics where discoveries are often claimed when the respective p-value is in the order of $2.7 \cdot 10^{-7}$ or lower [e.g., [Lyons, 2013](#)]. Additionally, because of the complexity of the models involved, for instance when the instrumental error must be taken in account, the data simulation that would be necessary to assess large significance requirements may be computationally prohibitive, or simply inefficient. Finally, corrections for multiple hypothesis testing may be of limited use because they are overly conservative [e.g., [Bonferroni, 1935, 1936](#)], require independence among the tests being conducted [e.g., [Hochberg, 1988](#)], or do not account for the enormous cost associated with a type I error [e.g., [Benjamini and Hochberg, 1995](#)].

In this thesis, we show how these challenges can be overcome by considering what is known in statistical literature as “*testing statistical hypotheses when a nuisance parameter is present only under the alternative*” or equivalently, “*testing with an unidentifiable parameter under H_0* ”. Specifically, we are interested in situations

where the model specified under the alternative hypothesis is characterized by the parameter being tested, hereafter $\boldsymbol{\eta}$, and an additional parameter, hereafter $\boldsymbol{\theta}$, which is not considered in the test and has no meaning under H_0 . This problem occurs for instance when testing the number of components that exist within a finite mixture model [e.g., Davison, 2003, p. 144]. Here we provide a high-level overview of existing methods and those we propose to tackle this class of problems. A more complete literature review appears in Chapter 1.

The general problem has long been studied, starting at least from the seminal work of Hotelling [1939], who introduced the so-called volume-of-tube formulae to assess the significance of the parameters in nonlinear regression models, and Davies [1977, 1987], where a stochastic process indexed by the nuisance parameter is considered, and the p-value of interest is obtained by approximating/bounding the tail probability of the supremum of the process via extreme value theory (EVT). Further works in the econometrics literature include Andrews and Ploberger [1994], who discuss optimal tests in presence of non-identifiable parameters under the alternative, and Hansen [1991, 1992, 1996], whose approach relies on the theory of empirical processes. In their practical implementation, all these inferential solutions reduce the problem of testing with an unidentifiable parameter under H_0 into one of *Testing One Hypothesis Multiple times* (TOHM). Specifically, a grid of size R over the parameter space of $\boldsymbol{\theta}$ is specified, and a single null hypothesis, H_0 , is tested against R different *sub-alternative hypotheses*, $H_{11}, \dots, H_{1r}, \dots, H_{1R}$, via R separate *sub-test statistics*, one for each value of the grid. It follows that each of the H_{1r} sub-alternatives is a special case of the *global alternative hypothesis* H_1 , where $\boldsymbol{\theta}$ is unknown. The number R of *sub-tests* is typically large, and the goal is to provide a global p-value as the standard of evidence for comparing H_0 and H_1 , i.e., the probability of rejecting H_0 when H_0 is true and $\boldsymbol{\theta}$ is unknown.

The above-mentioned methods often rely on case-by-case mathematical computations (e.g., Hotelling [1939], Davies [1977]), estimate integrals via total variations which can easily diverge (e.g., Davies [1987]), require the estimation of the covariance structure (e.g., Hansen [1991]), imply the specification of adequate weighting functions (e.g., Andrews and Ploberger [1994]), or rely on the full simulation of the empirical process involved (e.g., Hansen [1992, 1996]). Additionally, they may be limited to the one dimensional case, i.e., the non-identifiable parameter is assumed to be a scalar (e.g., Davies [1977, 1987]).

In this work we first discuss a computational efficient method to perform TOHM which overcomes these limitations in the one dimensional setting, i.e., when $\boldsymbol{\theta} =$

$\theta \in \Theta \subset \mathbb{R}$. The solution proposed relies on the theoretical framework of [Davies \[1977, 1987\]](#); specifically, we consider a stochastic process indexed by θ and we bound the distribution of its supremum considering the *expected number of upcrossings* of the process. Additionally, on the basis of a recent result proposed in the physics literature by [Gross and Vitells \[2010\]](#), we combine the probabilistic constructs of EVT with the flexibility of Monte Carlo simulation [[Algeri and van Dyk, 2017](#)]. This avoids the need of mathematical computations on a case-by-case basis, and drastically reduces the number of simulations required by a full simulation, especially under stringent significance requirements.

In practical applications, researchers often wish to avoid to use Bonferroni’s correction because of its perceived conservativeness. In this thesis we use classical EVT results for random sequences to identify situations where Bonferroni is equivalent to TOHM, and thus it can be used without being overly conservative [[Algeri and van Dyk, 2017](#)]. Further, in order to cover for situations where equivalence between the two approaches does not hold, we provide general guidelines to practitioners to select among different inferential procedures [[Algeri et al., 2016b](#)].

Finally, we extend [Algeri and van Dyk \[2017\]](#), and implicitly [Davies \[1977, 1987\]](#), to the multidimensional setting. In practice, we provide a simple, efficient and comprehensive tool to perform bump-hunting in two or more dimensions and to tackle other problems where structural changes can be characterized by a multidimensional parameter, $\boldsymbol{\theta}$. Specifically, as described in [Algeri and van Dyk \[2018\]](#) we consider a random field indexed by the non-identifiable multi-dimensional parameter, $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^D$, with $D \geq 1$, and on the basis of the seminal work of [Worsley \[1994\]](#), [Taylor and Adler \[2003\]](#), [Adler and Taylor \[2009\]](#), [Taylor and Worsley \[2008\]](#) pertaining to the distribution of the suprema of random fields, we use the *mean Euler characteristic of the excursion set* of the random field to approximate the global p-value. Unfortunately, closed-form expressions for the quantities involved are not easy to obtain. Thus, we propose a simple estimation procedure for the constructs involved, and we introduce a novel algorithm, based on graph theory, which allows a simplified computation of the Euler characteristic in multiple dimensions.

The remainder of the thesis is organized as follows. In Chapter 1 we review several non-standard hypothesis tests, specifically, those investigated by [Chernoff \[1954\]](#), [Davies \[1977, 1987\]](#), [Gross and Vitells \[2010\]](#), [Pilla et al. \[2005\]](#), [Pilla and Loader \[2005\]](#) and classical Multiple Hypothesis testing to control for the family-wise type I error [[Bonferroni, 1935, 1936](#)]. The methods discussed in Chapter 1 set the ground for the developments of the subsequent chapters in which we collect our novel con-

tributions. Specifically, in Chapter 2 we discuss how the approaches of Chernoff [1954] and Gross and Vitells [2010] can be combined to test non-nested models. In Chapter 3 we define the framework for TOHM and we derive a computable upper bound for the tail probability of interest by generalizing the method proposed by Gross and Vitells [2010]. In Chapter 4 we discuss the combined impact of classical EVT and stringent significance requirements on both the bound proposed in Chapter 3 and the usual Bonferroni's bound. In Chapter 5 we compare the approach of Pilla et al. [2005], Pilla and Loader [2005] with the one investigated by Gross and Vitells [2010] and we discuss simple guidelines on how to select among these methods and multiple hypothesis testing. In Chapter 6 we introduce the multidimensional extension of TOHM. We conclude with a general discussion of our results and future works. Finally, in Appendix A we quantify the rate of convergence of the bounds and approximations discussed in Chapters 3 and 4.

1

BACKGROUND ON HYPOTHESIS TESTING UNDER NON-STANDARD REGULARITY CONDITIONS

In this chapter, we propose a comprehensive review on statistical inference under standard and non-standard regularity conditions. Specifically, we first list all the assumptions required by classical asymptotic theory to guarantee the limiting normal distribution of the Maximum Likelihood Estimate (MLE) and the limiting χ^2 distribution of the Likelihood Ratio Test (LRT). Second, we discuss how the absence of some of these regularity conditions may affect the LRT and describe existing solutions to obviate these limitations. Finally, we consider a classical signal detection problem to outline the use of volume-of-tube formulae and multiple hypothesis testing to compute global p-values.

1.1 THE LARGE-SAMPLES DISTRIBUTION OF THE MLE AND LRT

1.1.1 CLASSICAL REGULARITY CONDITIONS

Let Y be a random variable with probability density $h(y, \boldsymbol{\eta}, \theta)$, from which a random sample of observations y_1, \dots, y_n is drawn. To simplify the notation, let $\boldsymbol{\delta} = (\boldsymbol{\eta}, \theta)$,

denote with Δ its parameter space, and write $h(y, \boldsymbol{\delta}) = h(y, \boldsymbol{\eta}, \theta)$. We let $D^k(\boldsymbol{\delta})$ be the k -dimensional array of the k^{th} order of partial derivatives of $\log h(y, \boldsymbol{\delta})$, $\boldsymbol{\delta}^*$ be the true value of $\boldsymbol{\delta}$ under which y is generated, and $\mathbf{I}(\boldsymbol{\delta}) = -E[D_n^2(\boldsymbol{\delta})]$ be the Fisher information matrix. It can be shown (e.g., Serfling [2009] and Davison [2003, p. 140]) that the MLE of $\boldsymbol{\delta}$, $\hat{\boldsymbol{\delta}}$, exists and

$$\sqrt{n}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}^*) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\delta}^*)) \quad (1.1)$$

if Conditions 1.1.1 listed below hold.

Conditions 1.1.1.

- B1. $\boldsymbol{\delta}^*$ is in the interior of Δ ;*
- B2. Let $D^k(\boldsymbol{\delta})_{i_1, \dots, i_k}$ be the element of $D^k(\boldsymbol{\delta})$ in position i_1, \dots, i_k . Then, for any value in the support of Y , $D^k(\boldsymbol{\delta})_{i_1, \dots, i_k}$ exists for all i_1, \dots, i_k and $k = 1, 2, 3$.*
- B3. For each $\boldsymbol{\delta}^\dagger \in \boldsymbol{\delta}$, there exist functions $h_k(y)$, eventually depending on $\boldsymbol{\delta}^\dagger$, such that for $\boldsymbol{\delta}$ in a ball $B(\boldsymbol{\delta}^\dagger)$, $|D^k(\boldsymbol{\delta})_{i_1, \dots, i_k}| < h_k(y)$ for i_1, \dots, i_k , for all y , and with $k = 1, 2, 3$. Further, for all $\boldsymbol{\delta} \in B(\boldsymbol{\delta}^\dagger)$ and all y , we have $\int h_k(y) \partial y < \infty$ when $k = 1, 2$, and $E[h_3(y)] < \infty$.*
- B4. $\mathbf{I}(\boldsymbol{\delta})$ exists and is positive-definite for $\boldsymbol{\delta}^\dagger \in \Delta$.*
- C1. The support of Y does not depend on $\boldsymbol{\delta}$.*
- C2. The model $h(y, \boldsymbol{\delta})$ is identifiable, i.e., $h(y, \boldsymbol{\delta}) \neq h(y, \boldsymbol{\delta}')$, $\forall \boldsymbol{\delta}, \boldsymbol{\delta}' \in \Delta$.*

Conditions B1-B4 corresponds to the classical regularity conditions which guarantee the asymptotic normality of the MLE, whereas Conditions C1 and C2, are necessary conditions for B2 and B4, respectively. Given their technical nature, B2 and B4 can only be assessed once the derivatives of $\log h(y, \boldsymbol{\delta})$ have been computed up to the third order; conversely Conditions C1 and Condition C2 can be easily verified once the probability density function $h(y, \boldsymbol{\eta}, \theta)$ is specified. Among Conditions 1.1.1, B1, B4 and C2 often fail in practice and they are the main object of our focus in this chapter.

1.1.2 TESTING UNDER NON-STANDARD REGULARITY CONDITIONS

Let y_1, \dots, y_n be a random sample of observations of the random variable Y introduced in Section 1.1.1. Suppose we wish to conduct the two-sided hypothesis

test

$$H_0 : \boldsymbol{\eta} = \boldsymbol{\eta}_0 \quad \text{versus} \quad H_1 : \boldsymbol{\eta} \neq \boldsymbol{\eta}_0. \quad (1.2)$$

We let $\boldsymbol{\eta} \in \Xi$ index the hypotheses in that $\boldsymbol{\eta} = \boldsymbol{\eta}_0$ defines H_0 , Ξ denotes the parameter space of $\boldsymbol{\eta}$, and let θ be a parameter that only exists under H_1 , it has no value under H_0 , and its parameter space is $\Theta \equiv [\boldsymbol{\mathcal{L}}; \boldsymbol{\mathcal{U}}] \subset \mathbb{R}$.

Common tests of hypothesis for (1.2) such as LRT, z -tests and t -tests typically rely on the known asymptotic distribution of their respective test statistics which in turn depends on the asymptotic normal distribution of the MLE in (1.1). These large-sample properties are guaranteed under Conditions 1.1.1. The presence of a nuisance parameter θ , which is defined only under H_1 , corresponds to a failure of condition C2 and (consequently) B4. This assumption is needed to ensure non-singularity of the Fisher Information matrix and consistency of the MLE. In the absence of B4 and C2, the asymptotic normality of the MLE is not guaranteed under H_0 , and neither is the asymptotic distribution of test statistic which relies on this normality. The failure of condition C2 provides the main motivation of this work and allows us to set up our framework for TOHM.

We denote the likelihood function of $h(y, \boldsymbol{\eta}, \theta)$, by $\mathcal{L}_n(\boldsymbol{\eta}, \theta)$, i.e.,

$$\mathcal{L}_n(\boldsymbol{\eta}, \theta) = \prod_{i=1}^n h(y_i, \boldsymbol{\eta}, \theta).$$

and we let $\lambda_n(\theta)$ be the ratio of the likelihoods under H_0 and under H_1 with θ fixed, i.e.,

$$\lambda_n(\theta) = \frac{\mathcal{L}_n(\boldsymbol{\eta}_0, -)}{\mathcal{L}_n(\hat{\boldsymbol{\eta}}_\theta, \theta)} \quad (1.3)$$

where $\hat{\boldsymbol{\eta}}_\theta$ is the MLE of $\boldsymbol{\eta}$ under H_1 , with θ fixed. We can specify the LRT statistic for a given value of $\theta \in \Theta$, namely $T_n(\theta)$, as

$$T_n(\theta) = -2 \log \lambda_n(\theta). \quad (1.4)$$

In order to derive the asymptotic distribution of (1.4) we require Conditions 1.1.2, in addition to Conditions 1.1.1.

Conditions 1.1.2.

B5. Denote with Δ_0 and Δ_1 , the parameter space of $\boldsymbol{\delta}$ under H_0 and H_1 , respectively. Then, $\Delta_0 \subseteq \Delta$ and $\Delta_1 \equiv \Delta \setminus \Delta_0$.

C3. The models specified under H_0 and H_1 are nested.

Notice that C3 is necessary condition for B5. Throughout this chapter we assume that Conditions 1.1.2 are verified, whereas in Chapter 2 we discuss a simple approach to extend the LRT to situations where C3 and consequently B5 are not met.

From Wilks theorem [Wilks, 1938] it follows that, for a given value of $\theta \in \Theta$, Conditions 1.1.1-1.1.2 hold (including B4 and C2), and the LRT in (1.4) is such that

$$T_n(\theta) \sim \chi_s^2 \quad \text{as} \quad n \rightarrow \infty, \quad (1.5)$$

where s is the dimension of the parameter $\boldsymbol{\eta}$ being tested. We can then consider a stochastic process $\{T_n(\theta)\}$, indexed by $\theta \in \Theta$, whose components are defined as in (1.4) and such that, by (1.5), each has asymptotically marginal χ_s^2 distribution under H_0 . In general terms, knowing the asymptotic distribution of the marginals of $\{T_n(\theta)\}$ is not sufficient to determine the asymptotic distribution of the entire process. However, as discussed in Hansen [1991], this can be done if Conditions 1.1.3 are satisfied.

Conditions 1.1.3. Let $l_n(\boldsymbol{\eta}, \theta) = \log \mathcal{L}_n(\boldsymbol{\eta}, \theta)$ be the log-likelihood function, and write for brevity $l_n(\boldsymbol{\delta})$, with $\boldsymbol{\delta} \in \Delta$ defined as in Conditions 1.1.1. $S_n(\boldsymbol{\eta}, \theta) = \frac{\partial}{\partial \boldsymbol{\eta}} l_n(\boldsymbol{\eta}, \theta)$ is the score vector and $M_n(\boldsymbol{\eta}, \theta)$ is the matrix of elements $-\frac{\partial^2}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} l_n(\boldsymbol{\eta}, \theta)$. Let $\dot{\Xi}$ be some neighborhood of $\boldsymbol{\eta}_0$. We require that

- (i) Δ is compact;
- (ii) $l(\boldsymbol{\delta}) = \lim_{n \rightarrow \infty} E[l_n(\boldsymbol{\delta})]$ is continuous in $\boldsymbol{\delta}$ uniformly over Δ ;
- (iii) $l_n(\boldsymbol{\delta}) \xrightarrow{p} l(\boldsymbol{\delta})$ for all $\boldsymbol{\delta} \in \Delta$;
- (iv) $l_n(\boldsymbol{\delta}) - l(\boldsymbol{\delta})$ is stochastically equicontinuous in $\boldsymbol{\delta}$ over Δ ;
- (v) under H_0 , for all $\theta \in \Theta$, $l_n(\boldsymbol{\delta})$ is uniquely maximized over $\boldsymbol{\eta} \in \dot{\Xi}$ at $\boldsymbol{\eta}_0$;
- (vi) at $\boldsymbol{\eta} = \boldsymbol{\eta}_0$, $l_n(\boldsymbol{\delta})$ does not depend upon θ ;
- (vii) $M(\boldsymbol{\eta}, \theta) = \lim_{n \rightarrow \infty} E[M_n(\boldsymbol{\eta}, \theta)]$ and $V(\boldsymbol{\eta}, \theta) = \lim_{n \rightarrow \infty} nE[S_n(\boldsymbol{\eta}, \theta)S_n(\boldsymbol{\eta}, \theta)']$ are continuous in $(\boldsymbol{\eta}, \theta)$ over $\dot{\Xi} \times \Theta$;
- (viii) $[M_n(\boldsymbol{\eta}, \theta), V_n(\boldsymbol{\eta}, \theta)] \xrightarrow{p} [M(\boldsymbol{\eta}, \theta), V(\boldsymbol{\eta}, \theta)]$ for all $(\boldsymbol{\eta}, \theta)$ in $\dot{\Xi} \times \Theta$;
- (ix) $M_n(\boldsymbol{\eta}, \theta) - M(\boldsymbol{\eta}, \theta)$ and $V_n(\boldsymbol{\eta}, \theta) - V(\boldsymbol{\eta}, \theta)$ are stochastically equicontinuous in $(\boldsymbol{\eta}, \theta)$ over $\dot{\Xi} \times \Theta$.
- (x) $M(\boldsymbol{\eta}_0, \theta)$ and $V(\boldsymbol{\eta}_0, \theta)$ are positive definite uniformly over $\theta \in \Theta$.

Hansen [1991] shows that, if Conditions 1.1.3 hold and if $V(\boldsymbol{\eta}_0, \theta) = M^{-1}(\boldsymbol{\eta}_0, \theta)^a$, under H_0

$$\{T_n(\theta)\} \xrightarrow{d} \{T(\theta)\} \quad \text{as } n \rightarrow \infty \quad (1.6)$$

and

$$\sup_{\theta \in \Theta} \{T_n(\theta)\} \xrightarrow{d} \sup_{\theta \in \Theta} \{T(\theta)\} \quad \text{as } n \rightarrow \infty$$

where $\{T(\theta)\}$ is a χ_s^2 -process i.e., a process

$$\{T(\theta)\} = \sum_{j=1}^s \{Z_j(\theta)\}^2 \quad (1.7)$$

where $\{Z_1(\theta)\}, \dots, \{Z_s(\theta)\}$ form a vector Gaussian process with mean zero and unit variance. Hansen [1991] also shows that the covariance function of $\{T(\theta)\}$, i.e., $\rho^T(\theta, \theta^\dagger) = \text{cov}(T(\theta), T(\theta^\dagger))$ corresponds to

$$\rho^T(\theta, \theta^\dagger) = \lim_{n \rightarrow \infty} \text{cov}(T_n(\theta), T_n(\theta^\dagger)). \quad (1.8)$$

Thus, we can combine the local LRT statistics $T_n(\theta)$ into a global LRT statistic by considering their supremum, and an asymptotic *global p-value* for the test in (1.2) can be specified as in (1.9)

$$P\left(\sup_{\theta \in \Theta} \{T(\theta)\} > c\right), \quad (1.9)$$

where c is the observed value of $\sup_{\theta \in \Theta} \{T_n(\theta)\}$. Notice that, instead of $\sup_{\theta \in \Theta} \{T_n(\theta)\}$, a different global test statistic can be specified. Andrews and Ploberger [1994] for instance consider $E[\exp\{T_n(\theta)\}]$, or equivalently one could simply refer to $E[\{T_n(\theta)\}]$. The reason why in this thesis we focus on the supremum of the LRT-process is dual. First of all, recall that the classical generalized LRT specifies

$$T_n = -2 \log \frac{\mathcal{L}_n(\boldsymbol{\eta}_0, -)}{\mathcal{L}_n(\hat{\boldsymbol{\eta}}, \hat{\theta})},$$

where $\hat{\boldsymbol{\eta}}$ and $\hat{\theta}$ are the MLEs of $\boldsymbol{\eta}$ and θ , respectively. It is easy to see [Hansen, 1991, Theorem 3] that

$$T_n = \sup_{\theta \in \Theta} \{T_n(\theta)\}$$

and thus, the results to be discussed below provide a direct generalization of classical inferential methods based on the LRT. Second, referring to the supremum of $\{T_n(\theta)\}$,

^aThis is typically the case in absence of serial correlation and heteroskedasticity. Alternatively, if $V(\boldsymbol{\eta}_0, \theta) \neq M^{-1}(\boldsymbol{\eta}_0, \theta)$, $\{T_n(\theta)\}$ converges in distribution to a scaled χ^2 -process [Hansen, 1991].

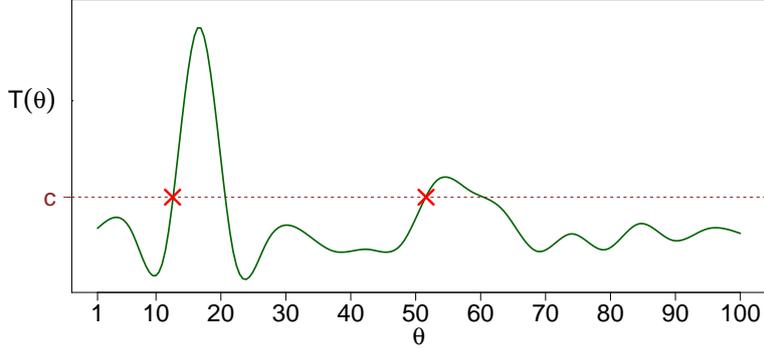


Figure 1.1: Upcrossings (red crosses) of the threshold c by the process $\{T(\theta)\}$.

instead of his mean for instance, allows us to conveniently exploit classical EVT results to bound/approximate probabilities of the form of (1.9), also known as an *excursion probabilities* [Adler, 2000].

Specifically, we can conveniently bound (1.9) using the concept of the upcrossings of a stochastic process (see Figure 1.1). We say that the process $\{T(\theta)\}$ has an *upcrossing* of a threshold $c \in \mathbb{R}$ at $\theta_0 \in \Theta$ if, for some $\epsilon > 0$, $\{T(\theta)\} \leq c$ in the interval $(\theta_0 - \epsilon, \theta_0)$ and $\{T(\theta)\} \geq c$ in the interval $[\theta_0, \theta_0 + \epsilon)$ [Adler, 2000]. Let N_c^T be the process of upcrossings of c by $\{T(\theta)\}$ over $\Theta \equiv [\mathfrak{L}, \mathfrak{U}]$. Following Cramér and Leadbetter [2013, p. 272] the p-value in (1.9) can be rewritten as

$$\begin{aligned} P\left(\sup_{\theta \in \Theta} \{T(\theta)\} > c\right) &= P(\{T(\mathfrak{L}) > c\} \cup \{N_c^T \geq 1\}) \\ &\leq P(T(\mathfrak{L}) > c) + P(N_c^T \geq 1) \end{aligned} \quad (1.10)$$

where $P(T(\mathfrak{L}) > c)$ is the probability that the process $\{T(\theta)\}$ exceeds c at the lower bound of the parameter space of θ , i.e., at $\theta = \mathfrak{L}$. Finally, (1.11) follows from the last line of (1.10) by Markov's inequality

$$P\left(\sup_{\theta \in \Theta} \{T(\theta)\} > c\right) \leq P(T(\mathfrak{L}) > c) + E[N_c^T]. \quad (1.11)$$

Because each component of $\{T(\theta)\}$ is χ_s^2 distributed, $P(T(\mathfrak{L}) > c) = P(\chi_s^2 > c)$; whereas, $E[N_c^T]$ is the expected number of upcrossings of c by $\{T(\theta)\}$ over the search region Θ and which bounds the probability $P(N_c^T \geq 1)$.

Search	Degree of surprise	Impact	Significance	P-value
Single top	None	Low	3σ	$< 1.35 \cdot 10^{-3}$
Neutrino oscillations	Medium	Medium	4σ	$< 3.16 \cdot 10^{-5}$
Higgs boson	Medium	Very high	5σ	$< 2.87 \cdot 10^{-7}$
Dark matter	Medium	High	5σ	$< 2.87 \cdot 10^{-7}$
4 th generation neutrinos	Substantial	High	6σ	$< 9.87 \cdot 10^{-10}$
Gravitational waves	None	High	7σ	$< 1.28 \cdot 10^{-12}$
Pentaquark	Substantial	High/very high	7σ	$< 1.28 \cdot 10^{-12}$
Faster-than-light neutrino	Enormous	Enormous	$8\sigma+$	$< 6.66 \cdot 10^{-16}$

Table 1.1: From Lyons [2013]. Some searches for new astrophysical phenomena, “degree of surprise”, impact of the discovery, proposed significance levels, σ -significance, and p-values for claiming a discovery.

Davies [1987] computes (1.11) as

$$P\left(\sup_{\theta \in \Theta} \{T(\theta)\} > c\right) \leq P(T(\mathfrak{L}) > c) + \frac{c^{\frac{s-1}{2}} e^{-\frac{c}{2}}}{2^{\frac{s}{2}} \Gamma(\frac{s}{2})} \int_{\mathfrak{L}}^{\mathfrak{U}} \left| \frac{\partial \sqrt{\{T(\theta)\}}}{\partial \theta} \right| d\theta, \quad (1.12)$$

Similar results are discussed in both Davies [1977] and Davies [1987] for the normal case. In Davies [1987], the main requirements for $\{T(\theta)\}$ are continuity of the process itself and its first derivative, except possibly for a finite number of jumps in the derivative. These assumptions guarantee the desired smoothness of the process and ensure that $E[N_c^T] < \infty$. Further Davies [1987] notices that (1.11) becomes sharp as $c \rightarrow \infty$ and if Condition 1.1.4 holds on $\rho^T(\theta, \theta^\dagger)$.

Condition 1.1.4. $\rho(\theta, \theta^\dagger) \rightarrow 0$ as $|\theta - \theta^\dagger| \rightarrow \infty$.

Condition 1.1.4 is known as *long-range independence* [Falk et al., 2010, p. 361]; if it is met, N_c^T assumes Poisson character as $c \rightarrow \infty$, and thus $P(N_c^T \geq 1) \approx E[N_c^T]$. Hence the sharpness of the bound in (1.11).

Davies [1987] estimates the integral in (1.12) via total variation. Unfortunately, as pointed out in Hansen [1991], there exist situations where the number of jumps in the derivative of $\{T(\theta)\}$ may cause the total variation to diverge. An alternative solution can be used to overcome this problem and has had significant impact in the physics community. Namely, Gross and Vitells [2010] introduces a novel method to correct for the so-called “look-elsewhere effect”, i.e., the reduction in significance that occurs

when one “looks” at many tests in the search for a significant result. This method has had a wide range of applications in various searches for new physics including in the discovery of the Higgs boson [Della Negra et al., 2012, van Dyk, 2014, Chatrchyan et al., 2012, Aad et al., 2012]. The approach of Gross and Vitells [2010] drastically reduces the size of the simulation required for a Monte Carlo evaluation of the null distribution of the LRT, and shows how a Monte Carlo estimate of $E[N_c^T]$ can be obtained using a more manageable simulation size. This is particularly advantageous in searches in high energy physics for instance, where the significance necessary to claim a discovery is in the order of at least $3\sigma^b$ (see Table 1.1), and the corresponding significance threshold c is typically very large; hence, upcrossings of c are expected to occur infrequently when simulating under H_0 . Gross and Vitells [2010] reduce the computational effort by estimating the expected number of upcrossings for a smaller threshold $c_0 \ll c$, and adapt (1.11) by rewriting it in terms of $E[N_{c_0}^T]$,

$$P\left(\sup_{\theta \in \Theta} \{T(\theta)\} > c\right) \leq P(T(\mathfrak{L}) > c) + \left(\frac{c}{c_0}\right)^{\frac{s-1}{2}} e^{-\frac{c-c_0}{2}} E[N_{c_0}^T]. \quad (1.13)$$

Gross and Vitells [2010] do not provide a formal argument to justify (1.13) and limit their attention to the LRT case and the resulting χ_s^2 -process. In Chapter 3, we formally derive (1.13), we extend it to a more general class of tests, and we discuss efficient choices of c_0 . In Chapter 6 we further extend these results to the case of θ being a multidimensional parameter.

In practical applications, H_0 may lie on the boundary of the parameter space, i.e., condition B1 may fail. Consider for instance an astrophysical signal search where the density of the events observed over the energy spectrum can be modelled as a mixture, i.e.,

$$(1 - \eta)f(y, \phi) + \eta g(y, \theta) \quad (1.14)$$

where $f(y; \phi)$ and $g(y, \theta)$ are the density functions of the background (known astrophysics) and signal (what we want to discover) respectively, y is the detected energy, ϕ and θ are nuisance parameters, and $0 \leq \eta \leq 1$ is the proportion of signal counts.

^bNotice that in physics the evidence in support of the new signal is often reported in terms of σ -significance, i.e., the number of standard deviations from the mean of a standard normal distribution that corresponds to the tail probability expressed by the one-sided p-value,

$$\#\sigma = \Phi^{-1}(1 - \text{p-value}),$$

where Φ is the standard normal cumulative function.

A test of hypothesis to assess the presence of the signal can be specified as

$$H_0 : \eta = 0 \quad \text{versus} \quad H_1 : \eta > 0 \quad (1.15)$$

and the LRT, for each θ fixed, can be written as in (1.16)

$$T_n(\theta) = -2 \sum_{i=1}^n \left\{ \log f(y_i, \hat{\phi}_0) - \log \left[(1 - \hat{\eta}_\theta) f(y_i, \hat{\phi}_{1\theta}) + \hat{\eta}_\theta g(y_i, \theta) \right] \right\} \quad (1.16)$$

where $\hat{\phi}_0$ is the MLE of ϕ under H_0 , $\hat{\eta}_\theta$ and $\hat{\phi}_{1\theta}$ are the MLEs of η and ϕ under H_1 , respectively, with θ fixed. If the value of θ is known (i.e., C2 and B4 among Conditions 1.1.1 hold), Chernoff [1954] shows that in this scenario the limiting null distribution of $T_n(\theta)$ in (1.16) is not χ_1^2 , but is a 50:50 mixture of χ_1^2 and zero, and known in the literature as $\bar{\chi}_{01}^2$ distribution [Lin and Lindsay, 1997, Takemura and Kuriki, 1997]. In Chapter 2 we discuss how this result can be combined with the solution of Gross and Vitells [2010] to tackle situations where H_0 is on the boundary of the parameter space, i.e., B1, B4 and C2 among Conditions 1.1.2 fail.

1.2 VOLUME-OF-TUBE FORMULAE

As an alternative to the upcrossings heuristic of Davies [1987] and Gross and Vitells [2010], methods relying on the so-called volume-of-tube formulae (e.g., Hotelling [1939], Pilla et al. [2005], Pilla and Loader [2005]) can be used to approximate the tail probabilities of the supremum of Gaussian related random processes. The use of the volume-of-tube formulae has received particular attention in recent physics literature [Pilla et al., 2005] and is purported to be more powerful than the usual LRT based approaches. Unfortunately, the mathematical implementation of the method is not straightforward, which strongly limits its use within the physics community. In this review, we outline this approach in the context of the signal detection problem introduced at the end of Section 1.1.2; readers are directed to Pilla et al. [2005], Pilla and Loader [2005], Adler [2000] and Adler and Taylor [2009] for further details and extensions of this approach to the multidimensional setting.

In Pilla et al. [2005] and Pilla and Loader [2005], the sub-test statistic considered is the normalized Score function. For the model in (1.14), when testing (1.15), this

statistics specifies as

$$S_n^*(0, \hat{\phi}_0, \theta) = \sum_{i=1}^n \frac{\left[\frac{g(y_i, \theta)}{f(y_i, \hat{\phi}_0)} - 1 \right]}{\sqrt{n\rho^*(\theta, \theta)}} \quad (1.17)$$

for all $\theta \in \Theta$, and $\rho^*(\theta, \theta^\dagger)$ is given by

$$\rho^*(\theta, \theta^\dagger) = \rho^{\mathfrak{s}}(\theta, \theta^\dagger) - \frac{\rho(\theta|\hat{\phi}_0)\rho(\theta^\dagger|\hat{\phi}_0)}{I(\hat{\phi}_0)}, \quad (1.18)$$

where $\hat{\phi}_0$ is the MLE of ϕ under H_0 , $I(\hat{\phi}_0) = \int \frac{\partial^2 \log f(y, \phi)}{\partial^2 \phi} \Big|_{\phi=\hat{\phi}_0} \partial y$ is the Fisher information evaluated at $\hat{\phi}_0$, $\rho(\theta|\hat{\phi}_0) = \int g(y, \theta) \frac{\partial \log f(y, \phi)}{\partial \phi} \Big|_{\phi=\hat{\phi}_0} \partial y$, and

$$\rho^{\mathfrak{s}}(\theta, \theta^\dagger) = \int \frac{g(y, \theta)g(y, \theta^\dagger)}{f(y, \hat{\phi}_0)} \partial y - 1 \quad (1.19)$$

is the covariance function of the unnormalized Score function $S(0, \hat{\phi}_0, \theta) = \frac{g(y, \theta)}{f(y, \hat{\phi}_0)} - 1$, and specifies

$$\rho^{\mathfrak{s}}(\theta, \theta^\dagger) = \text{cov} \left(S(0, \hat{\phi}_0, \theta), S(0, \hat{\phi}_0, \theta^\dagger) \right).$$

The random process of interest is the normalized Score process $\{S_n^*(\theta)\} = \{S_n^*(0, \hat{\phi}_0, \theta)\}$ and similarly to Section 1.4 the goal is to conveniently approximate the global p-value in (1.20)

$$P \left(\sup_{\theta \in \Theta} \{S_n^*(\theta)\} > c \right). \quad (1.20)$$

In order to guarantee validity of the results of Pilla et al. [2005] and Pilla and Loader [2005] described below, we specify the following Conditions 1.2.1.

Conditions 1.2.1.

- (i) Θ is a compact and convex subset of \mathbb{R} ;
- (ii) $\rho^{\mathfrak{s}}(\theta, \theta) < \infty$ for all $\theta \in \Theta$;
- (iii) Let \mathcal{Y}_f be the support of y when y has density function $f(y, \phi)$, and \mathcal{Y}_g be the support of y when its density is given by $g(y, \theta)$. Then, for each $\theta \in \Theta$, $\mathcal{Y}_g \subset \mathcal{Y}_f$.
- (iv) For all $y \in \mathcal{Y}_f$, $g(y, \theta)$ is twice differentiable and

$$\int \frac{[\frac{\partial}{\partial \theta} g(y, \theta)]^2}{f(y, \phi)} dy < \infty \quad \int \frac{[\frac{\partial^2}{\partial^2 \theta} g(y, \theta)]^2}{f(y, \phi)} dy < \infty;$$

(v) $\rho^s(\theta, \theta)$ is positive in θ and equivalent $f(y, \phi) \neq g(y, \theta)$ for all $\theta \in \Theta$;

(vi) $\rho^*(\theta, \theta^\dagger)$ is continuous and $0 < \rho^*(\theta, \theta) < \infty$ for all $\theta \in \Theta$.

Pilla and Loader [2005] show that, under (ii)-(vi) among Conditions 1.2.1,

$$\lim_{n \rightarrow \infty} P\left(\sup_{\theta \in \Theta} \{S_n^*(\theta)\} > c\right) = P\left(\sup_{\theta \in \Theta} \{Z(\theta)\} > c\right) \quad (1.21)$$

where $\{Z(\theta)\}$ is a zero-mean Gaussian process with covariance function given by

$$\rho^Z(\theta, \theta^\dagger) = \frac{\rho^*(\theta, \theta^\dagger)}{\sqrt{\rho^*(\theta, \theta)\rho^*(\theta^\dagger, \theta^\dagger)}}. \quad (1.22)$$

The problem of approximating $P(\sup_{\theta \in \Theta} \{Z(\theta)\} > c)$ is then converted into one of determining the volume of a manifold as described below.

Consider the uniformly convergent Karhunen-Loève expansion (e.g., Adler [1981]) of $\{Z(\theta)\}$ in (1.23), up to the J^{th} term

$$\{Z_J(\theta)\} = \left\{ \sum_{k=1}^J w_k(\theta) Z_k \right\} \quad (1.23)$$

where Z_k , $k = 1, \dots, J$ are independent and identically distributed standard Gaussian random variables, $\{w_k(\theta)\}_{k=1}^J$ is a sequence of twice continuously differentiable functions such that $\sum_{k=1}^J w_k^2(\theta) = 1$, and J is large enough to guarantee that $\{Z(\theta)\}$ is well approximated by $\{Z_J(\theta)\}$. We can then write

$$P\left(\sup_{\theta \in \Theta} \{Z(\theta)\} > c\right) \approx P\left(\sup_{\theta \in \Theta} \left\{ \sum_{k=1}^J w_k(\theta) Z_k \right\} > c\right) \quad (1.24)$$

$$= P\left(\sup_{\theta \in \Theta} \left\{ \sum_{k=1}^J \frac{Z_k w_k(\theta)}{\sqrt{\sum_{k=1}^J Z_k^2}} \right\} > \frac{c}{\sqrt{\sum_{k=1}^J Z_k^2}}\right) \quad (1.25)$$

$$= \int_{c^2}^{\infty} P\left(\sup_{\theta \in \Theta} \left\{ \sum_{k=1}^J w_k(\theta) U_k \right\} > \frac{c}{\sqrt{x}}\right) h_{\chi_J}(x) \partial x \quad (1.26)$$

where $(U_1, \dots, U_J) = \left(\frac{Z_1}{\sqrt{\sum_{k=1}^J Z_k^2}}, \dots, \frac{Z_J}{\sqrt{\sum_{k=1}^J Z_k^2}} \right)$ is a uniform random vector, $\mathbf{w}(\theta) = (w_1(\theta), \dots, w_J(\theta))$ is a curve over the unit sphere \mathcal{S}^{J-1} embedded in \mathbb{R}^J , and $h_{\chi_J}(x)$ is the probability density function of a χ_J^2 random variable. Notice that by Cauchy-

Schwartz inequality

$$\left| \sum_{k=1}^J w_k(\theta) U_k \right| \leq \sqrt{\sum_{k=1}^J w_k^2(\theta)} \sqrt{\sum_{k=1}^J U_k^2} \leq 1$$

and consequently $\frac{c}{\sqrt{x}} \leq 1$ and $c^2 < x < \infty$. Finally, we have

$$\sum_{k=1}^J (U_k - w_k(\theta))^2 = 2 \left(1 - \sum_{k=1}^J U_k w_k(\theta) \right)$$

and thus for any c_U such that $\sum_{k=1}^J U_k w_k(\theta) \geq c_U$ we write $\sqrt{\sum_{k=1}^J (U_k - w_k(\theta))^2} \leq \sqrt{2(1 - c_U)}$. Therefore,

$$P \left(\sup_{\theta \in \Theta} \left\{ \sum_{k=1}^J w_k(\theta) U_k \right\} \geq c_U \right) = P \left(\inf_{\theta \in \Theta} \left\{ \sqrt{\sum_{k=1}^J (U_k - w_k(\theta))^2} \right\} \leq \sqrt{2(1 - c_U)} \right) \quad (1.27)$$

$$= \frac{\mathcal{V}(\sqrt{2(1 - c_U)}, \mathcal{M})}{2\pi^{\frac{J}{2}} / \Gamma(\frac{J}{2})} \quad (1.28)$$

where $\mathcal{V}(\sqrt{2(1 - c_U)}, \mathcal{M})$ is the volume of the tube of radius $\sqrt{2(1 - c_U)}$ around the curve $\mathcal{M} = \{\mathbf{w}(\theta) : \theta \in \Theta\}$ and the denominator in the right hand side of (1.27) corresponds to the volume of \mathcal{S}^{J-1} . Thus, (1.28) follows from (1.27) since (U_1, \dots, U_J) is uniformly distributed over \mathcal{S}^{J-1} .

Several authors have derived expressions for $\mathcal{V}(\sqrt{2(1 - c_U)}, \mathcal{M})$ under suitable regularity conditions [e.g., [Hotelling, 1939](#), [Weyl, 1939](#), [Knowles and Siegmund, 1989](#), [Sun, 1993](#)], and these expressions typically involve geometric constants depending of the geometry of \mathcal{M} . Under (ii)-(vi) among Conditions 1.2.1, [Pilla and Loader \[2005\]](#) specify explicit closed-form expressions for the geometric constants involved and write

$$P \left(\sup_{\theta \in \Theta} \{Z(\theta)\} > c \right) = \frac{\kappa_0}{2\pi} P(\chi_2^2 \geq c^2) + \frac{P(\chi_1^2 \geq c^2)}{2} + o\left(\frac{e^{-\frac{c^2}{2}}}{c}\right) \quad (1.29)$$

where

$$\kappa_0 = \int \sqrt{\frac{\partial^2}{\partial \theta \partial \theta^\dagger} \rho^Z(\theta, \theta^\dagger)} \Big|_{\theta^\dagger = \theta} \partial \theta. \quad (1.30)$$

Thus, the global p-value is computed via (1.29) and, given the complexity of (1.22),

(1.30) is typically calculated via numerical integration.

Notice that the tests statistic proposed by Pilla et al. [2005] and Pilla and Loader [2005] and the one proposed by Gross and Vitells [2010] are asymptotically equivalent for large sample sizes. Similarly, approximating excursion probabilities via the volume-of-tube method or the upcrossings heuristic lead to the same result when c is sufficiently large. In Chapter 5, we further investigate these aspects by comparing the two approaches in terms of power and false discovery rate when dealing with small sample sizes, and we illustrate the difficulties associated with their implementation.

1.3 DETECTING SIGNALS VIA MULTIPLE HYPOTHESIS TESTING

In scenarios where we expect at most one signal or one structural change to be observed, Multiple Hypothesis Testing (MHT) can also be performed. In MHT several tests are conducted simultaneously, any of which can result in a false detection. While the individual tests are designed to control their specific false detection rate, namely α , the overall probability of having at least one false detection, i.e., the family-wise type I error, namely α_F , increases as the number of tests increases, leading to a higher rate of false discoveries than the one observed in each individual test.

In the framework introduced in Section 1.1.2, where a signal search is conducted over a region Θ , MHT is performed by fixing a set of values $\Theta_R = \{\theta_1, \dots, \theta_R\}$ over Θ and for each $\theta_r \in \Theta_R$ a test of hypothesis is performed. Consider for instance using the LRT in (1.16) as the test statistic to test (1.15), an ensemble of R local p-values of the form $P(T(\theta_r) > t(\theta_r))$ is produced, where $t(\theta_r)$ is the observed value of $T(\theta_r)$, and the smallest p-value, i.e.,

$$p_L = \min_{\theta_r \in \Theta_R} P(T(\mathfrak{L}) \geq t(\theta_r)) \quad (1.31)$$

is selected and corrected afterwards in order to guarantee the desired family-wise type I error rate (e.g., Efron [2012]).

If the R tests are independent under H_0 , e.g., if detecting a signal in a given c location does not depend on detections (or non-detections) in other locations, it can be easily shown that a global p-value, namely p_G , can be obtained via (1.32) (e.g.,

Kuehl [1994])

$$p_G = 1 - (1 - p_L)^R. \quad (1.32)$$

Consider a toy example in which we have 50 grid points over the search range Θ and 50 independent tests are conducted at the 5σ significance level, i.e., $\alpha = 1 - \Phi(5) = 2.87 \cdot 10^{-7}$ for each test. The chance of having at least one false detection among the 50 tests, i.e., the family-wise type I error rate, namely α_F , is

$$\alpha_F = 1 - (1 - \alpha)^R = 1.4 \cdot 10^{-5} \quad (1.33)$$

which corresponds to 4.18σ significance. This is approximately 50 times larger than the established probability of type I error associated with 5σ for each individual test, hence the need of correcting p_L according to (1.32).

Conversely, if the R tests are not independent of one another, controlling for the false detection rate is more difficult. In this scenario, the specific dependence structure among the R tests varies on a case-by-case basis and thus, contrary to (1.33), an exact general relationship between α and α_F , cannot be established. However, by means of Boole's inequality [e.g., Bonferroni, 1936] we have that

$$\alpha_F \leq R \cdot \alpha. \quad (1.34)$$

The adjusted global p-value corresponding to (1.34) is known as the Bonferroni's correction [Bonferroni, 1935, 1936] and specifies

$$p_{BF} = R \cdot p_L \quad (1.35)$$

which bounds p_G in (1.32) in that $p_G \leq p_{BF}$. In particular, p_{BF} is a first order approximation of p_G , i.e., expanding p_G via Taylor around zero we obtain

$$1 - (1 - p_L)^R = R \cdot p_L + O(p_L^2).$$

Thus p_G and p_{BF} are equivalent when dealing with strong signals, i.e., when $p_L \rightarrow 0$. This is reflected in the toy example above, where p_{BF} is equal to p_G and also leads to 4.18σ significance. (Recall $\frac{\alpha_F}{\alpha_L} \approx 50$ in the toy example.) Despite their easy implementation, the Bonferroni procedure is often dismissed by practitioners because, in addition to the stringent requirements to control for the overall false detection rate, it artificially depends on the number of tests R . This is particularly troublesome given the typically arbitrary nature of setting R when discretizing the

search region Θ .

Notice that the level at which the optimization and the correction occur differs between MHT and the methods in (1.13) and (1.29). Specifically, in p_{BF} the optimization is over the local p-values. In (1.13) and (1.29), on the other hand, the optimization is over the sub-test statistics, and an overall correction for the type I error is generated intrinsically by using the topology of underlying stochastic process to approximate the global p-value. We discuss in Chapter 4, however, practical situations in which p_{BF} approximates (1.13), and thus can be used without being overly conservative. In Chapter 5 we outline a simple sequential approach to select between TOHM and MHT while maintaining the desired statistical properties.

2

TESTING NON-NESTED MODELS VIA THE LIKELIHOOD RATIO TEST

The methods reviewed in Section 1.1.2 allow us to extend the LRT to test hypotheses conducted on the boundary of the parameter space and when a nuisance parameter is present only under the alternative. The goal of this chapter is to show how these solutions can be combined to select between non-nested models, a problem of crucial importance in physical discoveries and which received particular attention in the search for dark matter. Since the latter offers the main motivation of the approach discussed in this chapter, we briefly introduce the physics problem in Section 2.1. Applications to dark matter searches are discussed in Section 2.3.

2.1 NON-NESTED MODELS COMPARISON IN DARK MATTER SEARCHES

Dark matter was postulated in the 1930s by Jan Oort and Fritz Zwicky [Oort, 1932, Zwicky, 1933, 1937] to account for the orbital velocities of the stars within galaxies and to explain evidence of missing mass in the universe. Along with dark energy, it is hypothesized to constitute 95% of the universe. Understanding its nature and proving experimentally its existence is one of the most investigated problems in particle physics and astronomy.

Astrophysical searches for dark matter can be conducted via indirect detection, direct detection or colliders. In *indirect detection*, the search for dark matter is

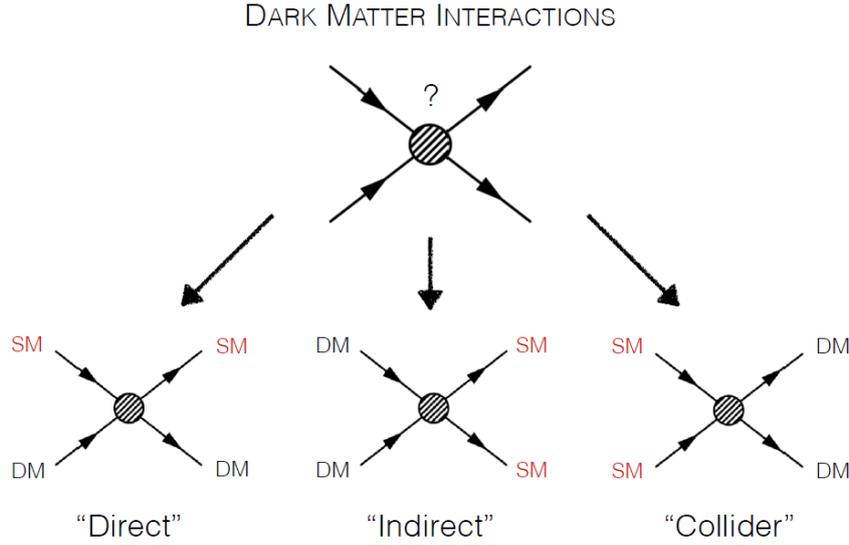


Figure 2.1: From [Kavanagh \[2018\]](#). The graphs above summarize the interactions of dark matter particles (DM) and elementary particles in the Standard Model (SM) which researchers aim to measure via direct detection, indirect detection and colliders. In direct detection the detector measures the recoil of DM particles with nuclei of SM particles. In indirect detection telescopes measure excesses of SM particles due to the annihilation of DM particles. Finally, colliders aim to produce DM particles via the collision of SM particles.

conducted by means of telescopes which aim to measure the products of dark matter particles annihilation, such as excesses of photons from dwarf galaxies and galaxy clusters, or emissions of charged particles (i.e., protons or electrons) from the Galactic center. Conversely, when conducting *direct detection*, the search for dark matter is done in underground mines with the attempt to measure the interaction of dark matter particles with nuclei of elementary particles. Finally, searches via *colliders* aim to produce dark matter particles via collisions of Standard Model particles^a [[Conrad, 2015](#), [Kavanagh, 2018](#)] (see Figure 2.1).

The data analyses proposed in this thesis mainly involve realistic simulated data of the Fermi Large Area Telescope (LAT). The Fermi LAT [[Atwood et al., 2009](#)] is a pair-conversion γ -ray telescope on board the Fermi satellite orbiting earth since 2008, and it produces images and measures energies associated to γ -rays between about a 100 MeV and several TeV.

In indirect searches, γ -rays are considered to be the “gold channel” in the discovery of dark matter, and a crucial step is to properly distinguish γ -rays emissions caused by conventional, known astrophysical sources and those originating from dark matter

^aThe Standard Model for particle physics is the model which describes all the fundamental particles and forces known today, as well as their interactions [e.g. [Ellis et al., 2012](#)].

annihilations. In this setting, the goal is not limited to assessing the presence of a signal in addition to the astrophysical background, but it also requires the ability to discriminate between γ -rays excesses due to dark matter emissions and those caused by conventional astrophysical sources mimicking them. In statistical terms, assessing the presence of a new emission translates into a signal detection problem (e.g., “bump-hunting”, change-point detection); whereas characterizing the astrophysical nature of the observed excess requires the ability of adequately select between non-nested models. Hence the need of suitable statistical procedures to address these tasks.

2.2 STATISTICAL COMPARISON OF NON-NESTED MODELS

The asymptotic χ^2 distribution of the LRT requires the models specified under H_0 and H_1 to be nested, that is the two densities to be compared are special cases of a full model and share the same parameter space. If they are not, C3 and B5, among Conditions 1.1.2, fail. For instance, let the models in (1.14), i.e., $f(y, \phi)$ and $g(y, \theta)$, being non-nested. The problem occurs when testing

$$H_0 : f(y, \phi) \quad \text{versus} \quad H_1 : g(y, \theta). \quad (2.1)$$

Notice that ϕ and θ are unknown parameters and thus the problem cannot be reduced to a test for simple hypotheses as in Cousins et al. [2005]. Specifically, Cousins et al. [2005] considers the case where $f(y, \phi)$ and $g(y, \theta)$ are fully specified and no unknown parameters are present. Thus, the LRT for the test in (2.1) reduces to a sum of independent and identically distributed random variables, and its asymptotic normality follows by the Central Limit Theorem (see Cox [1961] for more details). In this chapter, we limit our attention to the case where the unknown parameters ϕ and θ are one-dimensional; generalization to the multidimensional setting will be covered in Chapter 6.

Although $f(y, \phi)$ and $g(y, \theta)$ are non-nested, as shown in Algeri et al. [2016a], the framework of Section 1.1.2 can be extended to compare them by reformulating the comparison as a test of hypothesis with H_0 on the boundary, and in which a nuisance parameter is identified only under H_1 . Specifically, following the intuition of Cox [1962] and Atkinson [1970], we specify a comprehensive model that embeds $f(y, \phi)$ and $g(y, \theta)$. There are two reasonable formulations. We encountered the first in

(1.14); whereas the second one is

$$\frac{\{f(y, \phi)\}^{1-\eta}\{g(y, \theta)\}^\eta}{\int \{f(y, \phi)\}^{1-\eta}\{g(y, \theta)\}^\eta \partial \mathbf{y}}, \quad (2.2)$$

As discussed in Cox [1962, 2013], Atkinson [1970] and Quandt [1974], there are advantages and disadvantages to both. The additive form in (1.14) has the advantage of more appealing mathematical properties. Specifically, since no normalizing constant is involved, the maximization of the log-likelihood reduces to numerical optimization. Whereas (2.2) does not require η to be restricted to $0 \leq \eta \leq 1$ in order to obtain a valid probability density. In virtue of its computational efficiency, hereafter we focus on the formulation in (1.14). Similar results can be obtained when considering (2.2).

It is important to point out that in contrast to the test discussed in Section 1.1.2, here the model in (1.14) is not viewed as a mixture of astrophysical models in which a certain proportion of events, η , originates a process (e.g., the signal) represented by one model, and the remaining proportion, $1 - \eta$, originates from the competing process (e.g., the background) represented by the other model. Instead, here (1.14) is a mathematical formalization used to embed the densities $f(y, \phi)$ and $g(y, \theta)$ and η has no physical interpretation. Rather, as in Quandt [1974], η represents an auxiliary parameter which allows us to exploit the normality of its MLE to apply well-known asymptotic results. Specifically, this reduces the problem to a nested model comparison where the test of hypothesis is expressed as in (1.15).

Perhaps a more natural formulation of (1.2) would be $H_0 : \eta = 0$ versus $H_1 : \eta = 1$. Unfortunately, neither Wilks's or Chernoff's theorems apply to this formulation since they rely on the asymptotic normality of the MLE under H_0 , which can only hold if there is a continuum of possible values of η under H_1 , with $\eta = 0$ in its interior, and thus all of Conditions 1.1.1 and 1.1.2 fail. However, in addition to (1.15), as suggested in Cox [1962, 2013], the hypotheses

$$H_0 : \eta = 1 \quad \text{versus} \quad H_1 : \eta < 1 \quad (2.3)$$

should also be tested in order to exclude situations where η is within the $(0, 1)$ interval, leading to uninterpretable/uninformative results. In a similar spirit, Atkinson [1970] proposes to also test intermediate situations such as $H_0 : \eta = 0.5$ versus $H_1 : \eta \neq 0.5$.

The informative scenarios arising from (1.15) and (2.3) are the following:

- if H_0 in (1.15) is rejected and H_0 in (2.3) is not, $g(y, \theta)$ is selected,
- if H_0 in (2.3) is rejected and H_0 in (1.15) is not, $f(y, \phi)$ is selected.

In all other cases (1.15) and (2.3) are insufficient or inappropriate to select between the models being compared. Testing both (1.15) and (2.3) is particularly suited to particle physics searches where researchers assign different degrees of belief to the models being tested. Specifically, $f(y, \phi)$ typically corresponds to the physical background, i.e., everything which is already known and described in the Standard Model, whereas $g(y, \theta)$ is the density of the new emission, and main object of focus. As discussed in van Dyk [2014], the most stringent significance requirements (e.g., Table 1.1) are typically used only in the *detection* stage, i.e., when testing (1.15) to assess the presence of a new signal. Conversely, in the *exclusion* stage, i.e., when testing (2.3) to exclude the hypothesis of a signal being present, a significance level of 0.05 is typically sufficient.

Notice that for the model in (1.14) both the tests in (1.15) and (2.3) are conducted on the boundary of the parameter space of η . Additionally, in (1.15) θ is not identifiable under H_0 , and similarly ϕ is not identifiable when H_0 is true in (2.3), hence both C2 and B1 among Conditions 1.1.1 fail in this setting. However, we can combine the result of Chernoff [1954] and the results of Section 1.1.2 to overcome these difficulties. For simplicity, we now focus on the test (1.15); the same approach can be used when testing (2.3).

We consider the LRT-process, $\{T_n(\theta)\}$, introduced in Section 1.1.2. Under H_0 in (1.15), if Conditions B2-B4 among Conditions 1.1.1 and Conditions 1.1.3 hold, $\{T_n(\theta)\}$ converges in distribution, as $n \rightarrow \infty$, to a $\bar{\chi}_{01}^2$ -process, whose components are distributed as a mixture of χ_0^2 and χ_1^2 with weights 0.5 [Taylor and Worsley, 2013], and denoted by $\{K(\theta)\}$. In this setting, the global p-value, $P(\sup_{\theta \in \Theta} \{K(\theta)\} > c)$, can be written as

$$P\left(\sup_{\theta \in \Theta} \{K(\theta)\} > c\right) \leq P(K(\mathcal{L}) > c) + e^{-\frac{c-c_0}{2}} E[N_{c_0}^K]. \quad (2.4)$$

where, given that each component of $\{K(\theta)\}$ is $\bar{\chi}_{01}^2$ distributed, $P(K(\mathcal{L}) > c)$ corresponds to the probability that a χ_1^2 random variable is greater than c divided by 2. The correction term multiplying $E[N_{c_0}^K]$ is equivalent to the one in (1.13) with $s = 1$. This is not surprising since, for $c > 0$, $\{K(\theta)\}$ can only cross c when it assumes χ_1^2 character; however, we postpone the formal argument regarding the formal derivation of (2.4) to Chapter 3. The right hand side of (2.4) to bound/approximate the

global p-value of both tests in (1.15) and (2.3).

2.3 VALIDATION ON DARK MATTER MODELS AND APPLICATION TO SIMULATED DATA FROM THE FERMI-LAT

We illustrate the reliability of the method discussed in Section 2.2 for testing non-nested models using two examples relative to searches for dark matter.

Example 2.1. We let

$$f(y, \phi) \propto y^{-(\phi+1)} \quad (2.5)$$

be the probability density function of the γ -rays energies, denoted by y , originating from known cosmic sources mimicking dark matter emissions. Whereas the density function of the γ -ray energies associated with the true dark matter emission [Bergström et al., 1998] is given by

$$g(y, \theta) \propto 0.73 \left(\frac{y}{\theta}\right)^{-1.5} \exp\left\{-7.8\frac{y}{\theta}\right\} \quad (2.6)$$

with $\phi > 0$, $y \in [1, 100]$ and $\theta \in [1, 100]$. The goal is to decide which model between $f(y, \phi)$ and $g(y, \theta)$ provides a better fit.

Example 2.2. We make the same comparison of Example 2.1 but in the presence of power-law (Pareto Type I) distributed background. In this case, for the test in (2.1) H_0 specifies as

$$f(y, \zeta, \lambda, \phi) = (1 - \lambda) \frac{1}{k_\zeta y^{\zeta+1}} + \lambda \frac{1}{k_\phi y^{\phi+1}} \quad (2.7)$$

and H_1 specifies

$$g(y, \zeta, \lambda, \theta) = (1 - \lambda) \frac{1}{k_\zeta y^{\zeta+1}} + \lambda \frac{e^{-7.8\frac{y}{\theta}}}{y^{1.5} k_\theta}; \quad (2.8)$$

where k_ϕ , k_ζ and k_θ are the normalizing constants for each density, $0 < \lambda < 1$, $\zeta > 0$, $\phi > 0$, $y \in [1, 100]$ and $\theta \in [1, 100]$. Note the hierarchical nature of the model in (2.9) to be tested via (1.15) and (2.3)

$$(1 - \eta)f(y, \zeta, \lambda, \phi) + \eta g(y, \zeta, \lambda, \theta). \quad (2.9)$$

First, in both $f(y, \zeta, \lambda, \phi)$ and $g(y, \zeta, \lambda, \theta)$ the parameter λ , is used to specify the signal existence over a (relatively well known) background. Second, in (2.9) the

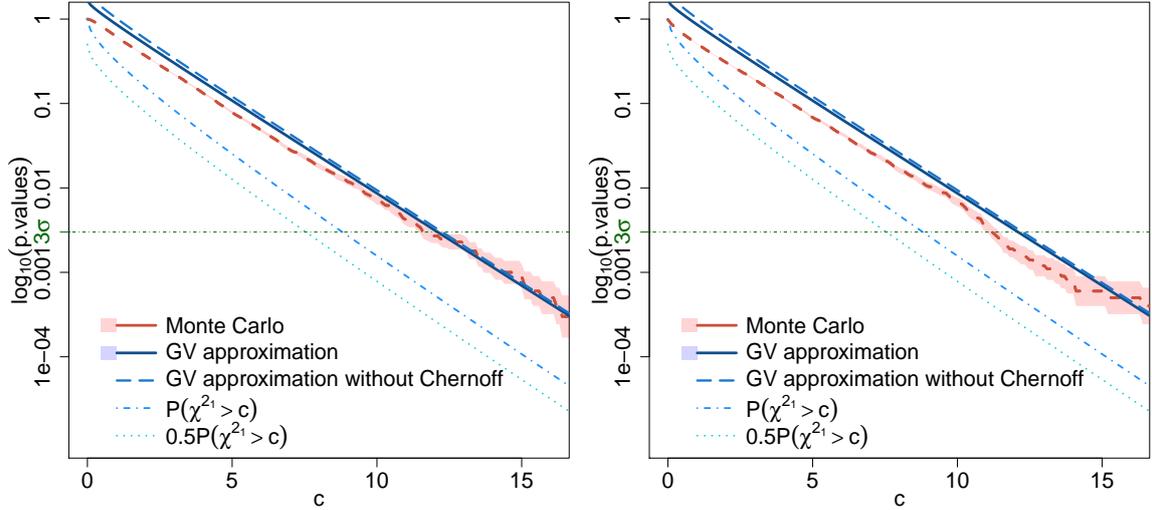


Figure 2.2: Comparing the bound/approximation in (2.4) (solid blue lines) with Monte Carlo estimates of $P(\sup_{\theta \in \Theta} \{K(\theta)\} > c)$ (red dashed lines), for Example 2.1 (left panel) and Example 2.2 (right panel). Approximations corresponding to (1.13) without the Chernoff correction (blue dashed lines), the χ^2 approximation $P(\chi_1^2 > c)$ (light blue dash-dotted lines) and the Chernoff-adjusted approximation $\frac{P(\chi_1^2 > c)}{2}$ (light blue dotted lines) are also reported. Monte Carlo p-values were obtained by simulating 10,000 datasets under H_0 , each of size 10,000 for both simulations. For each simulated dataset $\sup_{\theta \in \Theta} \{K(\theta)\}$ was computed over a grid of size 100 on $\Theta \equiv [1; 100]$ for Example 2.1 and size 400 for Example 2.2. Monte Carlo errors of the estimates of $P(\sup_{\theta \in \Theta} \{K(\theta)\} > c)$ are plotted in pink, whereas Monte Carlo errors associated with the Monte Carlo estimates of $E[N_{c_0}^K]$ in (2.4) are plotted in grey, but are too small to be visible.

parameter η is used as a merely mathematical tool to treat the non-nested case (as described in Section 2.2).

In both examples, we estimated the expected number of upcrossings $E[N_{c_0}^K]$ using 1,000 Monte Carlo replicates. The approximation to $P(\sup_{\theta \in \Theta} \{K(\theta)\} > c)$ is calculated using (2.4) on a grid of values of c . The results are compared with the respective Monte Carlo p-values in Figure 2.2 along with the nominal χ^2 and Chernoff corrections one might compute ignoring the failure of Conditions C2 and B4, among Conditions 1.1.1.

As expected, when c is small, the bound in (2.4) (blue solid lines) is greater than its Monte Carlo counterpart (red dashed lines). As c increases, however, the bound becomes sharp and converges to the Monte Carlo estimates for a good approximation of $P(\sup_{\theta \in \Theta} \{K(\theta)\} > c)$. When the failure of C2 and B4 among Conditions 1.1.1 is ignored, the p-values are computed assuming χ^2 or $\bar{\chi}_{01}^2$ distribution of the LRT (blue dotted-dashed lines and light blue dotted lines respectively). This approach leads

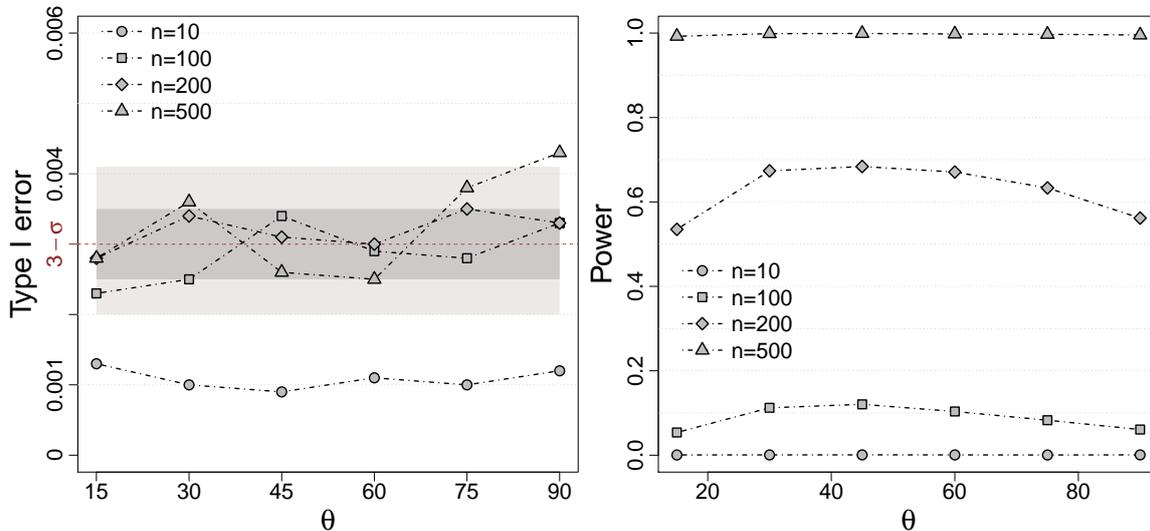


Figure 2.3: Simulated type I errors (left panel) and power functions (right panel) for Example 2.1 with at 3σ significance. Shaded areas indicate regions expected to contain 68% (dark gray) and 95% (light gray) of the symbols if the nominal type I error of 0.0013 holds. For both the type I error and power curves 10000 Monte Carlo simulations were used.

to over optimistic results, i.e., the p-values are always lower than those obtained via Monte Carlo. Finally, we compare the bound in (1.13), which ignores the failure of condition B1 among Conditions 1.1.1, with the bound in (2.4) which accounts for the fact that (1.15) is conducted on the boundary of the parameter space of η (blue dashed lines and blue solid line respectively). The two bounds lead to similar results. This is not surprising for two main reasons. First of all, the first terms in the right hand sides of (1.13) and (2.4) tend to zero as $c \rightarrow \infty$ and thus both bounds are dominated by the expected number of upcrossings of the process. Second, in the practical implementation of (1.13) and (2.4), an estimate for expected number of upcrossings is obtained via a Monte Carlo simulation of $\{T_n(\theta)\}$ under H_0 . The latter converges, for large sample sizes, to $\{K(\theta)\}$ and thus, even when ignoring that the test is conducted on the boundary, an estimate for $E[N_{c_0}^K]$ is automatically generated.

It is not uncommon in practice, and specifically in astrophysics, for the number of event counts to be considerably smaller than the 10,000 used in Figure 2.2. Thus, we conduct a simulation study to verify the type I error (i.e., the rate of false rejections of H_0) of the method with smaller samples and verify that the approximate p-value in (2.4) holds. The left panel of Figure 2.3 reports the simulated type I errors with

Example	H_0	n	$\hat{\eta}$	c	Significance
Example 2.1	$\eta = 0$	200	0.971	21.018	4.038σ
	$\eta = 1$	200	p-value = 0.528		
Example 2.2	$\eta = 0$	2726	0.999	12.096	2.673σ
	$\eta = 1$	2726	p-value > 1		

Table 2.1: Summary of the analysis on the Fermi LAT simulation comparing the models in Example 2.1 and Example 2.2. Estimates and significances refer to the tests $H_0 : \eta = 0$ versus $H_1 : \eta > 0$. P-values refer to the tests $H_0 : \eta = 1$ versus $H_1 : \eta < 1$.

a detection threshold on the p-value of 0.0013 (3σ) for different sample sizes for the models in Example 2.1. When the sample size is at least 100, the Monte Carlo results are consistent with the numerical 3σ error rate. The right panel of Figure 2.3 shows the power (probability of detection) curves at 3σ of the same test for different sample sizes. For all the values of θ considered (reported on the x-axis), a sample size of 500 is sufficient to achieve a power of nearly 1.

After having assessed that (2.4) provides a good approximation of the global p-value as c increases, we can now proceed with the application of our method to astrophysical data. Specifically, we consider simulated data from the Fermi Large Area Telescope (LAT) obtained with the *gtobssim* package^b, which include realistic representations of the effects of the detector and present backgrounds. We focus on a 5 years observation of dwarf galaxies and we introduce a putative dark matter emission at mass of $\theta=35$ GeV. This assumption is consistent with the most generic and popular models for dark matter, which assumes the latter to be made in large part made of Weakly Interacting Massive Particles (WIMP) [Conrad, 2015].

In Example 2.1, we assume that the astrophysical background, for example γ -rays introduced by our own Galaxy, was completely removed. Our simulation leads to about 200 events associated to the dark matter signal and detected by the Fermi LAT. In order to assess which model among (2.5) and (2.6) provides a better fit for the data, we perform both tests (1.15) and (2.3). As shown in Table 2.1, we find a 4.038σ significance in favor of the dark matter model.

Conversely, in Example 2.2, we consider the case where the background could not

^b<http://fermi.gsfc.nasa.gov/ssc/data/analysis/software>

be removed and thus we properly replace the models in (2.5) and (2.6) with those in (2.7) and (2.8) respectively, to account for this additional level of complexity. We then considered 2176 counts from a power-law distributed background source, and about 550 dark matter events. For simplicity, the mixture parameter λ is fixed at 0.2. In this case, we find 2.673σ significance in favor of the model in (2.8). As expected, introducing background significantly reduces the power for distinguishing a dark matter source from a conventional source. It should be noted however that (unlike in a full analysis) we do not attempt to remove the background by taking γ -ray directions into account.

The R package `NONnest` [Algeri, 2015] provides a software solution to implement the method described in this chapter and can be downloaded from <http://wwwf.imperial.ac.uk/~sa2514/Research.html>.

3

TESTING ONE HYPOTHESIS MULTIPLE TIMES: THE ONE-DIMENSIONAL CASE

The results of [Davies \[1987\]](#) reviewed in Chapter 2 provide a theoretical framework to conveniently bound/approximate the distribution of the supremum of χ^2 processes. Additionally, [Gross and Vitells \[2010\]](#) propose a computationally efficient procedure to compute the quantities involved when performing tests of hypothesis via the LRT and the underlying χ^2 -process. In this Chapter, we provide a formal justification for the testing procedure of [Gross and Vitells \[2010\]](#) and we generalize it beyond the LRT and the χ^2 case. This allows us to establish a generalized methodological framework to efficiently bound/approximate global p-values. As additional advantage, the resulting method only depends on the marginal distribution of the underlying process and thus generalizes beyond Conditions 1.1.3.

3.1 DEFINITION AND FORMALIZATION

Recall that $\theta \in \Theta \equiv [\mathcal{L}; \mathcal{U}]$ is the nuisance parameter which is defined only under H_1 . In order to generalize the notation of Chapter 2, we let $\{W(\theta)\}$ be a generic stochastic process indexed by $\theta \in \Theta$ with covariance function $\rho(\theta, \theta^t)$, and we let N_c be the process of upcrossings of a threshold $c \in \mathbb{R}$ by $\{W(\theta)\}$ over Θ . As in [Davies \[1987\]](#), we require $\{W(\theta)\}$ to be sufficiently smooth to guarantee that expected number of upcrossings $E[N_c]$ does not diverge. Hence, we stipulate Conditions 3.1.1.

Conditions 3.1.1.

- (i) $\{W(\theta)\}$ has continuous sample paths;
- (ii) $\{W(\theta)\}$ has continuous first derivative, except possibly for a finite number of jumps;
- (iii) the components $W(\theta)$ of $\{W(\theta)\}$ are identically distributed for all $\theta \in \Theta$.

Since (1.10) and (1.11) hold for any stochastic process [Cramér and Leadbetter, 2013, p. 272], we rewrite them more generally in (3.1), (3.2) and (3.3).

$$P\left(\sup_{\theta \in \Theta} \{W(\theta)\} > c\right) = P(\{W(\mathfrak{L}) > c\} \cup \{N_c \geq 1\}) \quad (3.1)$$

$$\leq P(W(\mathfrak{L}) > c) + P(N_c \geq 1) \quad (3.2)$$

where $P(W(\mathfrak{L}) > c)$ is the probability that the process $\{W(\theta)\}$ exceeds c at the lower bound of Θ ; whereas, as in (1.11), (3.3) follows from (3.2) by Markov's inequality

$$P\left(\sup_{\theta \in \Theta} \{W(\theta)\} > c\right) \leq P(W(\mathfrak{L}) > c) + E[N_c] \quad (3.3)$$

Our goal is to find a convenient way to estimate $E[N_c]$ and bound/approximate $P(\sup_{\theta \in \Theta} \{W(\theta)\} > c)$. Result 3.1.2 and Result 3.1.3 below are sufficient to achieve this goal.

Result 3.1.2. *Let $c \in \mathbb{R}$ be an arbitrary threshold, let $a(c)$ be a function which depends on c but not on θ , and $b(\Theta)$ be a function which does not depend on c , and to be calculated over the search region Θ . If $E[N_c]$ can be decomposed as*

$$E[N_c] = a(c)b(\Theta) \quad (3.4)$$

then,

$$E[N_c] = \frac{a(c)}{a(c_0)} E[N_{c_0}] \quad \forall c_0 \leq c, c_0 \in \mathbb{R}. \quad (3.5)$$

Proof. The proof is straightforward because the decomposition in (3.4) holds for any $c \in \mathbb{R}$, and thus also holds for any $0 < c_0 < c$, with $c_0 \in \mathbb{R}$. Consequently, (3.5) is obtained by solving the following system of linear equations

$$\begin{cases} E[N_c] = a(c)b(\Theta) \\ E[N_{c_0}] = a(c_0)b(\Theta). \end{cases}$$

□

In general, the derivation of a closed-form expression of $b(\Theta)$ in (3.4) may be challenging, and it typically requires knowledge of $\rho(\theta, \theta^\dagger)$. Therefore, (3.5) offers a significant advantage in computing $E[N_c]$, provided that, as discussed in Section 3.2, $E[N_{c_0}]$ can be estimated accurately.

The following result follows from (3.3), (3.4), and (3.5).

Result 3.1.3. *Under Conditions 3.1.1, the tail probability $P(\sup_{\theta \in \Theta} \{W(\theta)\} > c)$ can be bounded by*

$$P\left(\sup_{\theta \in \Theta} \{W(\theta)\} > c\right) \leq P(W(\mathfrak{L}) > c) + \frac{a(c)}{a(c_0)} E[N_{c_0}] \quad \forall c_0 \leq c, c_0 \in \mathbb{R}. \quad (3.6)$$

If additionally, Condition 1.1.4 holds, the bound in (3.6) approaches equality as $c \rightarrow \infty$.

Proof. Equation (3.6) follows from (3.3), (3.4), and (3.5). Additionally, under Conditions 3.1.1 and Condition 1.1.4, we expect N_c to assume Poisson character as $c \rightarrow \infty$ [Leadbetter et al., 1983, Davies, 1977], and thus

$$P(N_c > 1) \approx 1 - e^{-E[N_c]}. \quad (3.7)$$

Consequently the right hand side of (3.7) is well approximated by $E[N_c]$ as $c \rightarrow \infty$, and thus the right hand side of (3.3) is a sharp bound for (3.2). Finally, since the probability of the event $\{W(\mathfrak{L}) > c\} \cap \{N_c \geq 1\}$ is dominated by $P(N_c > 1)$, the bound in (3.2) also approaches equality. □

3.1.1 TOHM BOUNDS FOR GAUSSIAN-RELATED PROCESSES

We now derive explicit forms of (3.3) for several Gaussian related processes including Gaussian, χ_s^2 , $\bar{\chi}_{01}^2$, F and t -processes.

GAUSSIAN PROCESS. Let $\{Z(\theta)\}$ be a mean zero and variance one Gaussian process, such that $Z(\theta) \sim N(0, 1)$ for all $\theta \in \Theta$, and let N_c^Z be the process of upcrossings of c by $\{Z(\theta)\}$ over $\Theta \equiv [\mathfrak{L}; \mathfrak{U}]$. Following Leadbetter et al. [1983, p. 150], we can write

$$E[N_c^Z] = E\left[\frac{\partial\{Z(\theta)\}}{\partial\theta} \mathbb{1}_{\frac{\partial\{Z(\theta)\}}{\partial\theta} > 0} \Big| \{Z(\theta)\} = c\right] f_z(c). \quad (3.8)$$

where $f_z(c)$ is the density function of a standard normal evaluated at c . Equation (3.8) implies that the expected number of upcrossings is given by the average of positive slopes of the sample functions of $\{Z(\theta)\}$ at c , multiplied by the density of $Z(\theta)$ at c . Thus, condition (3.4) is verified for the Gaussian case with

$$a(c) = e^{-\frac{c^2}{2}} \quad \text{and} \quad b(\Theta) \propto E \left[\frac{\partial \{Z(\theta)\}}{\partial \theta} \mathbb{1}_{\frac{\partial \{Z(\theta)\}}{\partial \theta} > 0} \middle| \{Z(\theta)\} = c \right]. \quad (3.9)$$

The bound in (3.10) follows from Results 3.1.2 and 3.1.3

$$P \left(\sup_{\theta \in \Theta} \{Z(\theta)\} > c \right) \leq \Phi(-c) + e^{-\frac{c^2 - c_0^2}{2}} E[N_{c_0}^Z]. \quad (3.10)$$

where $\Phi(-c)$ is the cumulative density function of a standard normal evaluated at $-c$ and the correction term $e^{-\frac{c^2 - c_0^2}{2}}$ is given by the ratio $\frac{a(c)}{a(c_0)}$, with $a(c)$ calculated as in (3.9). For the stationary case, the same result can be obtained by expressing $E[N_c^Z]$ considering Rice's formula [Rice, 1944] i.e.,

$$E[N_c^Z] = \frac{|\mathfrak{L} - \mathfrak{U}|}{2\pi} \sqrt{\rho''(\theta, \theta)} e^{-\frac{c^2}{2}}$$

where $\rho''(\theta, \theta) = \frac{\partial \theta}{\partial \theta \partial \theta^\dagger} \rho(\theta, \theta^\dagger) \Big|_{\theta=\theta^\dagger}$ is the second spectral moment of $\{Z(\theta)\}$ and is assumed to be finite, whereas $|\mathfrak{L} - \mathfrak{U}|$ is the length of Θ . As discussed in Davies [1987], for a two-sided test, the excursion probability of interest is $P(\sup_{\theta \in \Theta} |\{Z(\theta)\}| > c)$; the bound of which is twice the right hand side of (3.10).

χ^2 -PROCESS. As in Chapter 2, let $\{T(\theta)\}$ be a χ_s^2 -process with s degrees of freedom and such that $T(\theta) \sim \chi_s^2$ for all $\theta \in \Theta$. Let $E[N_c^T]$ be the expected number of upcrossings of c by $\{T(\theta)\}$ over Θ . On the basis of Sharpe [1978, p. 375], we can express $E[N_c^T]$ in the form

$$E[N_c^T] = E \left[\frac{\partial \{T(\theta)\}}{\partial \theta} \mathbb{1}_{\frac{\partial \{T(\theta)\}}{\partial \theta} > 0} \middle| \{T(\theta)\} = c \right] f_{\chi_s^2}(c). \quad (3.11)$$

Thus, condition (3.4) is verified for the χ_s^2 case with

$$a(c) = c^{\frac{s-1}{2}} e^{-\frac{c}{2}} \quad \text{and} \quad b(\Theta) \propto E \left[\frac{\partial \{T(\theta)\}}{\partial \theta} \mathbb{1}_{\frac{\partial \{T(\theta)\}}{\partial \theta} > 0} \middle| \{T(\theta)\} = c \right],$$

hence (1.13) follows from (3.5), with $P(T(\mathfrak{L}) > c)$ in (1.13) being the probability that a χ_s^2 is greater than c .

$\bar{\chi}_{01}^2$ - PROCESS. Let $\{K(\theta)\}$ be the $\bar{\chi}_{01}^2$ process introduced in Chapter 2 and such that $K(\theta) \sim \bar{\chi}_{01}^2$ for all $\theta \in \Theta$. Let N_c^K be the process of upcrossings of a threshold c by $\{K(\theta)\}$ over Θ . As in (3.11) we write

$$E[N_c^K] = E \left[\frac{\partial \{K(\theta)\}}{\partial \theta} \mathbb{1}_{\frac{\partial \{K(\theta)\}}{\partial \theta} > 0} \middle| \{K(\theta)\} = c \right] \frac{f_{\chi_1^2}(c)}{2} \quad (3.12)$$

with c typically non-negative and where $f_{\chi_1^2}(c)$ is the density function of a χ_1^2 random variable evaluated at c . Thus

$$a(c) = \frac{e^{-\frac{c}{2}}}{2^{\frac{3}{2}} \sqrt{\pi}} \quad \text{and} \quad b(\Theta) = E \left[\frac{\partial \{K(\theta)\}}{\partial \theta} \mathbb{1}_{\frac{\partial \{K(\theta)\}}{\partial \theta} > 0} \middle| \{K(\theta)\} = c \right]$$

from which (2.4) follows.

The rate of convergence of the difference between the right and left hand side of (1.13), (2.4), and (3.10) are discussed in Section A.1 of the Appendix A. We will further study the sharpness of the bounds in (1.13), (2.4) and (3.10), as $c \rightarrow \infty$ in Section 3.2 via a suite of simulation studies.

The bounds in (1.13), (2.4), (3.10), and the analogous for other Gaussian related processes such as F and t -processes, can also be derived on the basis of random fields theory. This, along with the multidimensional version of our bound for TOHM is the subject of Chapter 6 where, following [Algeri and van Dyk, 2018], it is shown that $a(c)$ in (3.6) is proportional to the *Euler characteristic density* of first order of the respective stochastic process evaluated at c (to be introduced in Chapter 6). As a result, the function $a(c)$ in (3.6) only depends on the marginal distribution of the components of the process $\{W(\theta)\}$. This introduces an additional advantage of using (3.6) to bound $P(\sup_{\theta \in \Theta} \{W(\theta)\} > c)$. Specifically, since (3.6) only require knowledge of the marginal distribution of the components of $\{W(\theta)\}$, the uniformity Conditions 1.1.3 are no longer needed to guarantee uniform convergence of $\{T_n(\theta)\}$ to $\{T(\theta)\}$ (or $\{K(\theta)\}$).

Explicit closed-form expressions for the Euler characteristic densities are well-known in random field theory for Gaussian-related processes [Worsley, 1994, Taylor and Worsley, 2008]. On the basis of these results, we report explicit forms of (3.6) for F and t processes below.

F -PROCESS. Consider a F -process $\{F(\theta)\}$ with s and v degrees of freedom and such that $F(\theta) \sim F_{s,v}$ for all $\theta \in \Theta$. Let $E[N_{c_0}^F]$ be the expected number of upcrossings

of c by $\{F(\theta)\}$, then the TOHM bound in equation (3.6) specifies as

$$P\left(\sup_{\theta \in \Theta} \{F(\theta)\} > c\right) \leq P(F(\mathfrak{L}) > c) + \left(\frac{c}{c_0}\right)^{\frac{s-1}{2}} \left(\frac{v + s \cdot c}{v + s \cdot c_0}\right)^{-\frac{s+v-2}{2}} E[N_{c_0}^F] \quad (3.13)$$

for all $c_0 \leq c, c_0 \in \mathbb{R}$, and with $a(c) = c^{\frac{s-1}{2}} (v + s \cdot c)^{-\frac{s+v-2}{2}}$.

t-PROCESS. Consider a *t*-process $\{V(\theta)\}$ with s degrees of freedom and such that $V(\mathfrak{L}) \sim t_s$. Let $E[N_{c_0}^V]$ be the expected number of upcrossings of c by $\{V(\theta)\}$, then the TOHM bound in equation (3.6) specifies as

$$P\left(\sup_{\theta \in \Theta} \{V(\theta)\} > c\right) \leq P(V(\mathfrak{L}) > c) + \left(\frac{1 + c^2}{1 + c_0^2}\right)^{-\frac{s-1}{2}} E[N_{c_0}^V] \quad (3.14)$$

for all $c_0 \leq c, c_0 \in \mathbb{R}$, and with $a(c) = (1 + c^2)^{-\frac{s-1}{2}}$.

The reader is referred to Chapter 6 and [Algeri and van Dyk \[2018\]](#) for more details on this approach.

3.1.2 TESTING ONE HYPOTHESIS MULTIPLE TIMES IN PRACTICE

In practice, we consider a sub-test statistic, $W_n(\theta)$, whose asymptotic or exact distribution under H_0 is known to be the same as a random variable $W(\theta)$, for all $\theta \in \Theta$, and we evaluate the stochastic process $\{W(\theta)\}$ on a finite set of values of θ . Specifically, let Θ_R represent a fine grid of evaluation points for $\{W(\theta)\}$, with R being the, typically large, resolution of Θ_R so that $\Theta_R = \{\theta_1, \dots, \theta_r, \dots, \theta_R\} \subseteq \Theta$. Let $\{W(\theta_r)\}$ be the random sequence which coincides with $\{W(\theta)\}$ at each grid point θ_r and $\{w(\theta_r)\}$ be the observed value of $\{W(\theta_r)\}$ at θ_r . We approximate the global test statistics $\sup_{\theta \in \Theta} \{W(\theta)\}$ with its discrete counterpart $\max_{\theta_r \in \Theta_R} \{W(\theta_r)\}$, and its observed value is given by

$$c_R = \max_{\theta_r \in \Theta_R} \{w(\theta_r)\} \quad (3.15)$$

and it is such that $c = \lim_{R \rightarrow \infty} c_R$. We let the upcrossings of c_R by $\{W(\theta_r)\}$ over Θ_R , namely \tilde{N}_{c_R} , be events of the type $\{W(\theta_{r-1}) \leq c_R, W(\theta_r) > c_R\}$. In order to ensure that Θ_R is sufficiently dense to guarantee that the process of upcrossings and the supremum of $\{W(\theta)\}$ are well approximated by the process of upcrossings and the maximum of $\{W(\theta_r)\}$, we require

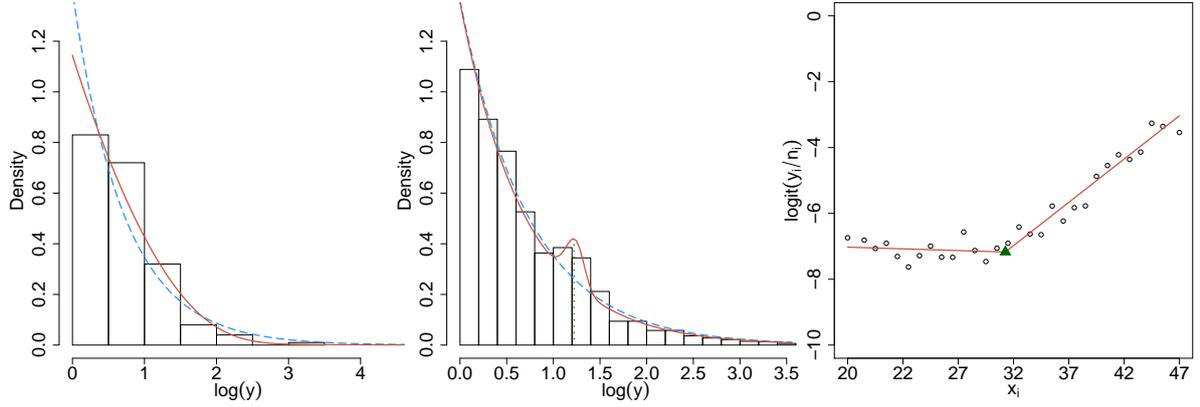


Figure 3.1: Left panel: histogram of the Fermi-LAT realistic data simulation for Example 2.1 (log-scale), null model (blue dashed curve) fitted as a power-law distributed cosmic source ($\hat{\phi}_0 = 1.395$) and fitted alternative model (red solid curve) with $\hat{\eta}_{\tilde{\theta}} = 0.971$, $\hat{\phi}_{1\tilde{\theta}} = 0.745$ and $\tilde{\theta} = 27.265$. Central panel: histogram of the Fermi-LAT realistic data simulation for Example 3.1 (on log-scale), null model (blue dashed curve) fitted under the assumption of background only counts ($\hat{\phi}_0 = 1.350$), and fitted alternative model (red solid curve) with $\hat{\eta}_{\tilde{\theta}} = 0.045$, $\hat{\phi}_{1\tilde{\theta}} = 1.406$. The green dotted vertical line indicates the location of the observed Gaussian bump, i.e., $\tilde{\theta} = 3.404$. Right panel: Down syndrome data and fitted regression model (red piecewise-linear solid lines), with break-point (green triangle) at $\tilde{\theta} = 31.266$.

Conditions 3.1.4. As $R \rightarrow \infty$

$$\max_{\theta_r \in \Theta_R} \{W(\theta_r)\} \xrightarrow{d} \sup_{\theta \in \Theta} \{W(\theta)\}; \quad (3.16)$$

$$\tilde{N}_{c_R} \xrightarrow{a.s.} N_{c_R}. \quad (3.17)$$

By dominated convergence, it follows from (3.17) that

$$E[\tilde{N}_{c_R}] \rightarrow E[N_{c_R}] \quad \text{as } R \rightarrow \infty.$$

Cramér and Leadbetter [2013, p. 63 and 195] prove that (3.16) and (3.17) hold when, given any sequence $\{q_R\}$ such that $q_R \rightarrow 0$ as $R \rightarrow \infty$, the grid points in Θ_R are of the form rq_R , with $(r-1)q_R \in \Theta$ and $rq_R \in \Theta$, for $r = 1, 2, \dots, R$. However, since in practice Θ_R may be determined by the experiment, in Section 3.2 we discuss simple graphical tools to assess Conditions 3.1.4 in applied settings.

The right hand side of (3.6) can therefore be approximated, as $R \rightarrow \infty$, by

$$P(W(\mathfrak{L}) > c_R) + \frac{a(c_R)}{a(c_0)} E[\tilde{N}_{c_0}] \quad \forall c_0 \leq c_R, c_0 \in \mathbb{R}. \quad (3.18)$$

and thus, under Condition 1.1.4, for large R and c_R , (3.18) is an asymptotically sharp bound for $P(\sup_{\theta \in \Theta} \{W(\theta)\} > c_R)$; whereas, $E[\tilde{N}_{c_0}]$ can be replaced by its Monte Carlo estimate, namely $E[\tilde{N}_{c_0}]$ (see Section 3.2).

Thus, the problem of testing (1.15) is reduced to testing H_0 versus R *sub-alternative hypotheses*, namely H_{1r} , i.e., *Testing One Hypothesis Multiple Times*. For the R *sub-tests*, H_0 versus H_{1r} , the observed *sub-test statistics* $\{w(\theta_1), \dots, w(\theta_R)\}$, realizations of $\{W(\theta)\}$, are combined into the global test statistic c_R and a bound/approximation for the global p-value is computed via (3.18).

3.2 PRACTICAL MATTERS

3.2.1 CASE STUDIES: DESCRIPTION

We now discuss in details the implementation of TOHM in the context of three case studies: the non-nested models comparison proposed in Example 2.1, a “bump hunting” problem, and a logistic model with a break point; hereafter, we refer to the last two as Examples 3.1 and 3.2, respectively. As for Example 2.1, data in Example 3.1 were generated using realistic simulations of the Fermi Large Area Telescope (LAT).

The set up for Example 2.1 is the same described in Chapter 2 and the goal is to distinguish between a dark matter emission and a cosmic source mimicking it. The comprehensive model in (1.14), which facilitates the non-nested model comparison, specifies

$$(1 - \eta) \frac{1}{k_\phi y^{\phi+1}} + \eta \frac{y^{-1.5}}{k_\theta} \exp\left\{-7.8 \frac{y}{\theta}\right\} \quad (3.19)$$

with $0 \leq \eta \leq 1$, and we test both (1.15) and (2.3).

Example 3.1. Conversely from Example 2.1, here we aim to properly distinguish between γ -ray signals induced by dark matter annihilations and those induced by the astrophysical background. In this setting intermediate values of η correspond to the intensity of the dark matter signal. Under the model in (3.20), events induced by dark matter are modeled as a narrow Gaussian bump with mean energy θ and the astrophysical background is distributed as a power-law with index ϕ ,

$$(1 - \eta) \frac{1}{k_\phi y^{\phi+1}} + \frac{\eta}{k_\theta} \exp\left\{-\frac{(y - \theta)^2}{0.02\theta^2}\right\} \quad \text{for } y \geq 1, \quad (3.20)$$

where k_ϕ and k_θ are normalizing constants, $y \geq 1$, $\phi > 0$, and $\theta \geq 1$. In our simulation, we set $\theta = 3.5\text{GeV}$, $\phi = 1.4$, and we consider the energy band $y \in [1; 35]$. This setup resulted in 64 dark matter events and 2274 background events. For more physics details, see [Algeri et al. \[2016b\]](#). Fermi-LAT datasets for Examples 2.1 and 3.1 are plotted in the first two panels of Figure 3.1.

Example 3.2. We consider a simple logistic-regression model where the logit is modelled as

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \phi_1 + \phi_2 x_i + \xi(x_i - \theta) \mathbb{1}_{\{x_i \geq \theta\}} \quad \forall i = 1, \dots, n \quad (3.21)$$

with $\pi_i = P(y_i = 1|x_i)$, $x_i \in \mathbb{R}$ is the covariate of interest, $\theta \in \mathbb{R}$ is the location of the unknown break-point, and $\mathbb{1}_{\{\cdot\}}$ is the indicator function. The model in (3.21) is applied to the *Down Syndrome dataset* available in the R package `segmented` [[Muggeo et al., 2008](#)]. The dataset records whether babies born to 354,880 women are affected by Down Syndrome. We use (3.21) to model the probability, π_i , that a woman of age x_i has a baby with down syndrome, where $x_i \in [17; 47]$, and we let $\theta \in [20; 44]$. The logit of the ratio between the number of down syndrome cases and number of births by age group is plotted in the right panel of Figure 3.1.

In Example 2.1 and 3.1 we use the LRT statistic, $T_n(\theta)$, introduced in (1.4) as the sub-test statistic. Since both tests are either of the form in (1.15) or in (2.3), we test on the boundary of the parameter space of η and thus the asymptotic distribution of $T_n(\theta)$ under H_0 is $\bar{\chi}_{01}^2$ for each $\theta \in \Theta$ fixed. In Example 3.2, we test

$$H_0 : \xi = 0 \quad \text{versus} \quad H_1 : \xi \neq 0 \quad (3.22)$$

via the signed-root of the LRT which specifies, for all $\theta \in \Theta$,

$$Q_n(\theta) = \text{sign}(\hat{\xi}_\theta) \sqrt{T_n(\theta)} \quad (3.23)$$

where $\hat{\xi}_\theta$ is the MLE of ξ under H_1 at θ fixed. Hence the sub-tests statistics are asymptotically normally distributed under H_0 [e.g., [Davies, 1977](#)].

3.2.2 THE CHOICES OF c_0 AND R

As described in Section 3.1.2, in the practical implementation of TOHM we specify a grid Θ_R over the parameter space Θ , with R being the number of times H_0 is tested versus the ensemble of sub-alternatives H_{11}, \dots, H_{1R} , one for each value θ_r in Θ_R . In

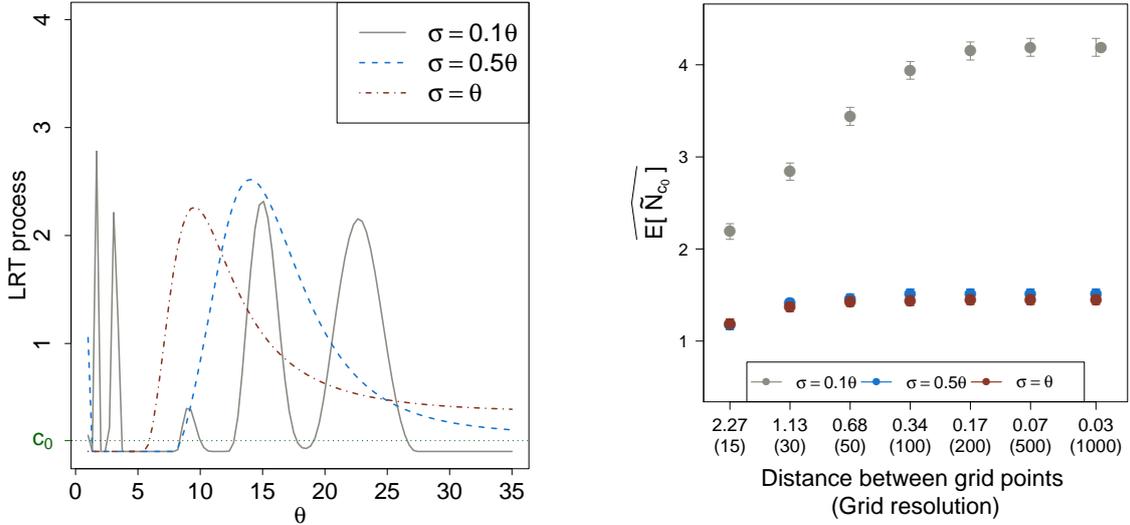


Figure 3.2: Left panel: simulated sample paths of the LRT process under H_0 in Example 3.1. Both plots consider different scaling factors of the width of the Gaussian bump, i.e., we set $\sigma = 0.1\theta, 0.5\theta$ and θ . Right panel: upcrossings plot showing Monte Carlo estimates of the expected number of upcrossings of $c_0 = 0.1$ by the LRT process under H_0 true, for Example 3.1, and evaluated over grids of resolutions $R = 15, 30, 50, 100, 200, 500$.

practice, R must either be chosen arbitrarily by the researcher or determined by the nature of the experiment. In either case, R must be sufficiently large to guarantee robustness of the results, yet small enough to ensure computational efficiency when calculating (3.18).

One possibility is to choose R large enough so that, for a given c_0 , the expected number of upcrossings of c_0 by $\{W(\theta_r)\}$, $E[\tilde{N}_{c_0}]$, converges to a finite limit, which we expect, by Conditions 3.1.4, to correspond to the expected number of upcrossings of c_0 by the underlying null continuous stochastic process $\{W(\theta)\}$, i.e., $E[N_{c_0}]$. This strategy requires us to set c_0 before setting R .

In Gross and Vitells [2010], c_0 is chosen to minimize the coefficient of variation $\frac{\sigma_{c_0}}{E[N_{c_0}]}$, where σ_{c_0} is the standard deviation of N_{c_0} . Under the assumption that the coefficient of variation decreases with increasing $E[N_{c_0}]$, the problem reduces to maximizing $E[N_{c_0}]$. From (3.4) it follows that $E[N_{c_0}] = a(c_0)b(\Theta)$, and thus we write

$$c_0 = \operatorname{argmax}_{c \in \mathbb{R}} a(c). \quad (3.24)$$

For instance, for the χ_s^2 process $\{T(\theta)\}$, $\hat{c}_0 = s - 1$ when $s > 1$. However, to cover cases where $c_0 = 0$, when the global maximum in (3.24) is not defined, or simply, as an alternative to (3.24), one can carry out a sensitivity analysis based on a Monte

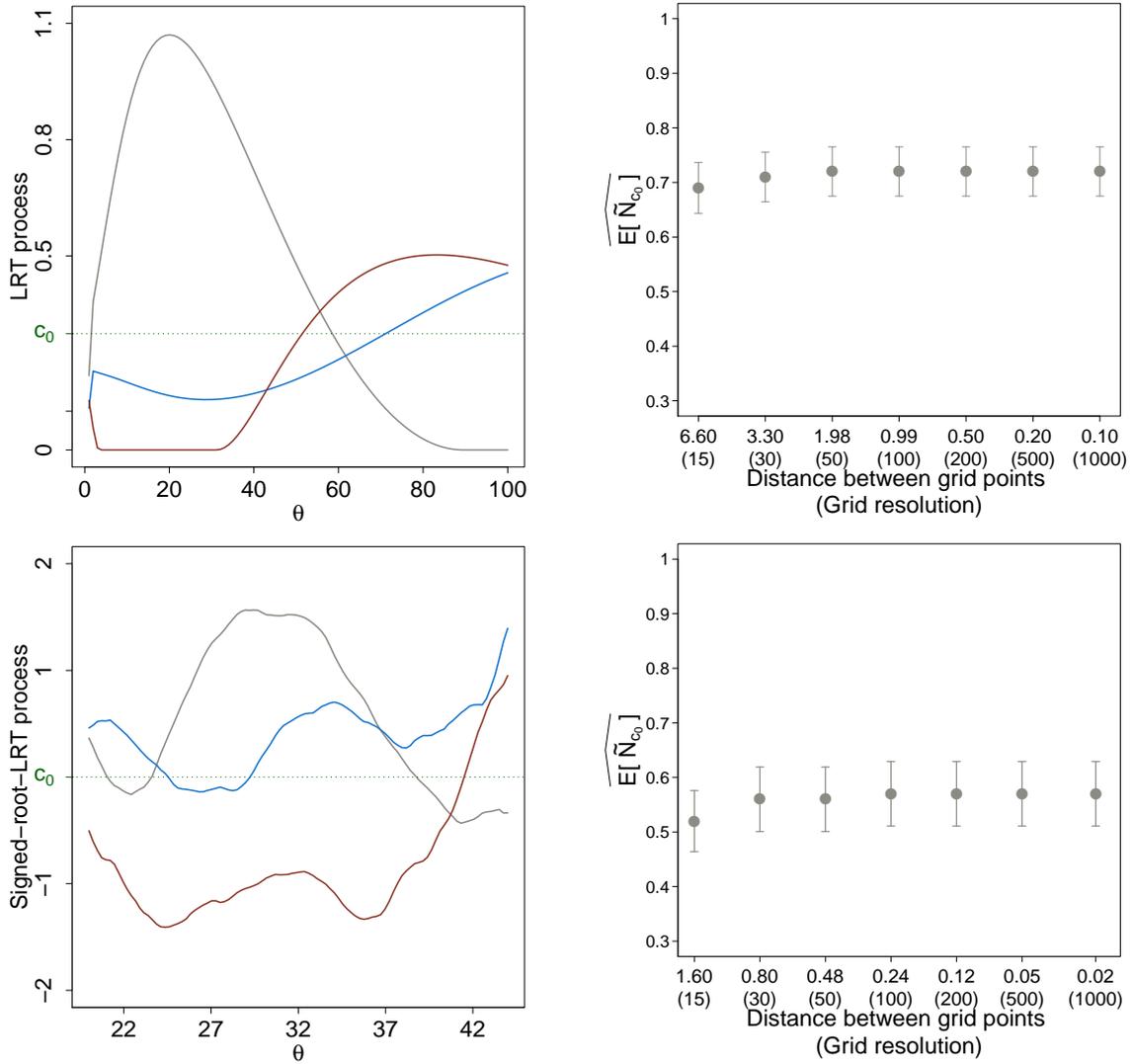


Figure 3.3: Left panels: simulated sample paths of the LRT process for Example 2.1 (upper left) and of the signed-root-LRT process for Example 3.2 (bottom left) considering three different random samples under H_0 . Right panels: upcrossings plots showing Monte Carlo estimates of the expected number of upcrossings under H_0 of $c_0 = 0.3$ (upper right) by the LRT process for Example 2.1 and of $c_0 = 0$ (bottom right) by the signed-root-LRT process for Example 3.2. In both cases we use grids of resolutions $R = 15, 30, 50, 100, 200, 500$.

Carlo simulation of the underlying processes under H_0 . This can be done as in the left panel of Figure 3.2 for Example 3.1, and the analogous for Examples 2.1 and 3.2 in the left panels of Figure 3.3. Looking at these plots, we choose c_0 to be the level (on the y-axis) with respect to which the process oscillates more often, and thus, with respect to which the upcrossings occur with higher frequency. This leads to values c_0 equal to 0.3, 0.1 and 0 for Examples 2.1, 3.1 and 3.2, respectively. Inspecting the smoothness of the trace plots in the left panels of Figures 3.2 and 3.3

also allows us to qualitatively assess Conditions 3.1.1 and 3.1.4, which are necessary for the validity of the results of Sections 3.1.

The *upcrossing plot* in the right panel of Figure 3.2 is a simple graphical tool that helps us to identify the value of R that best negotiates the trade-off between accuracy and computational efficiency discussed at the beginning of this section. This plot displays Monte Carlo estimates $\widehat{E[\tilde{N}_{c_0}]}$ for the LRT in Example 3.1, under H_0 , as a function of R , evaluated at $R = 15, 30, 50, 100, 200, 500, 1000$ equally spaced grid points over Θ . Analogous plots for Example 2.1 and 3.2 are reported in Figure 3.3. For each resolution we computed 100 Monte Carlo simulations. The sample size of each replicate must be reasonably large to guarantee the asymptotic distribution of the sub-test statistics, hence we choose a size of 1000 for each Monte Carlo replicate. In all our examples, 100 simulations are sufficient to achieve small Monte Carlo errors. If the number of upcrossings increases with R without converging, it means that the resolution is not sufficiently high to catch all the crossings or, the underlying process is not sufficiently smooth to guarantee $E[N_{c_0}] < \infty$. Conversely, if the number of upcrossings converges, as in the well-known scree-plot used for Principal Component Analysis (PCA) (e.g., James et al. [2013, p. 383]), we look for an “elbow” in the plot of $\widehat{E[\tilde{N}_{c_0}]}$. The value of R corresponding to the elbow is the smallest value for which $\widehat{E[\tilde{N}_{c_0}]}$ converges, by Conditions 3.1.4, to its limit $E[N_{c_0}]$ up to Monte Carlo error.

In the upcrossings plot in Figure 3.2 (right panel), we also investigate the relationship between the width of the signal in the bump-hunting example and the grid resolution. In particular, we replicate the simulation for three choices of the Gaussian width, namely $\sigma = 0.1\theta, \sigma = 0.5\theta$ and $\sigma = \theta$. (In our actual analysis $\sigma = 0.1\theta$.) The left panel in Figure 3.2 illustrates how the width of the signal affects the smoothness of the underlying processes. Note surprisingly, wider signals lead to smoother processes and consequently, $\widehat{E[\tilde{N}_{c_0}]}$ converges faster to its limit (right panel of Figure 3.2) and a lower grid resolution is often sufficient to identify all the upcrossings.

In general, the choice of R not only impacts the upper bound/approximation for the global p-value in (3.6), but also impacts the observed value of the test statistics, c_R , which we assume converges to c , as $R \rightarrow \infty$, see (3.15). Specifically, if the gap between θ_r and θ_{r+1} is wider than the signal width, c_R may underestimate c , and the signal may be missed. Thus, for signal identification problems where the signal is typically localized over a small region of the search interval $[\mathfrak{L}, \mathfrak{U}]$, a higher resolution is required not only to increase the accuracy of the estimate in (3.18), but

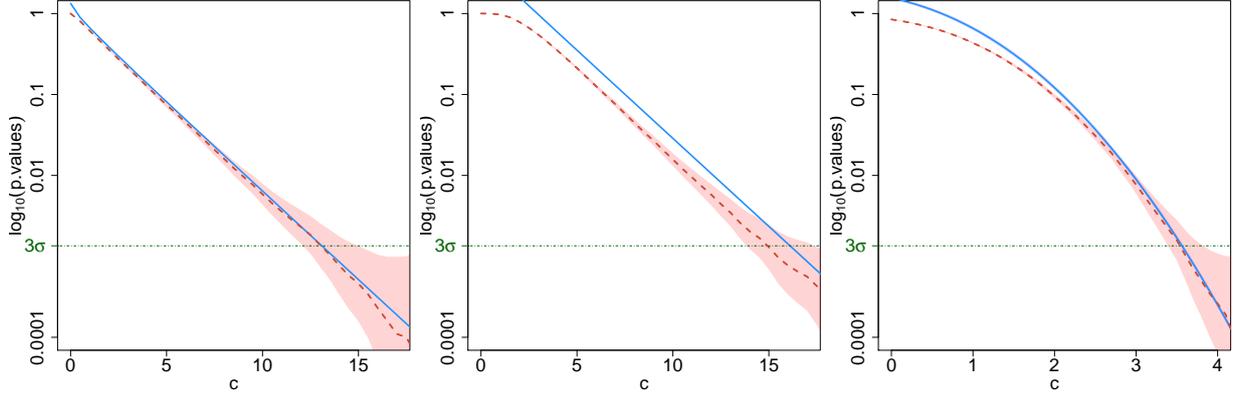


Figure 3.4: Estimated bound/approximation in (3.18) (blue solid line), simulated global p-values (on \log_{10} -scale), Monte Carlo estimates of $P(\sup_{\theta \in \Theta} \{W(\theta)\} > c)$ (red dashed line), and Monte Carlo Errors (pink areas) for increasing values of the threshold c , for Example 2.1 (left panel), Example 3.1 (central panel) and Example 3.2 (right panel). Monte Carlo errors associated with $\widehat{E[\tilde{N}_{c_0}]}$ on the bound in (3.18) are plotted in grey, but are too small to be visible.

also to avoid false negatives, which would in turn adversely affect the power of the test.

Conversely, in Examples 2.1 and 3.2, the signal is spread either over the whole parameter space or over a large portion of it. Thus, as shown in the right panels of Figure 3.3, convergence is achieved quickly because the underlying processes are characterized by smooth sample paths (left panels). In these cases the choice of R should be based on the level of accuracy of c_R as an estimate for the maximum of the underlying process, and its location over the search range, i.e.,

$$\tilde{\theta} = \operatorname{argmax}_{\theta_r \in \Theta_R} \{W(\theta_r)\}. \quad (3.25)$$

Finally, based on the elbow in the upcrossings plots, the values of R we select are $R = 50$ in Example 2.1, $R = 100$ in Example 3.1 (with $\sigma = 0.1\theta$ as in (3.20)), and $R = 30$ in Example 3.2. However, in order to guarantee accuracy of at least 0.5 for the selected location of the break-point, $\tilde{\theta}$, we set $R = 50$ in Example 3.2.

For each of the models considered, we computed (3.18) using the R and c_0 selected above. The results obtained are compared in Figure 3.4 with the Monte Carlo estimates of the global p-value, $P(\sup_{\theta \in \Theta} \{W(\theta)\} > c)$, for increasing values of c , obtained using 100,000 simulations, each of size 10,000. The pink areas correspond to the respective Monte Carlo errors. The Monte Carlo errors associated to the estimate $\widehat{E[\tilde{N}_{c_0}]}$ for $E[\tilde{N}_{c_0}]$ in (3.18) (and displayed on a lower scale in the upcrossings

plots) are also incorporated in Figure 3.4, but they are too small to be visible. As expected, the estimated TOHM bounds approach the “truth” (i.e., the Monte Carlo estimate of $P(\sup_{\theta \in \Theta} \{W(\theta)\} > c)$) as $c \rightarrow \infty$. Convergence appears to be slower for Example 3.1. The plots, however, are presented on \log_{10} -scale, and thus in all cases we obtain a good approximations of the global p-values.

4

TESTING ONE HYPOTHESIS MULTIPLE TIMES AND MULTIPLE HYPOTHESIS TESTING: TWO SIDES OF THE SAME COIN

The goal of this chapter is to highlight the connection between TOHM and Multiple Hypothesis Testing (MHT). Specifically, we use classical EVT results for random sequences to identify sufficient conditions under which, despite their difference in implementation, the classical Bonferroni's correction introduced in Section 1.3 can be used as reasonable approximation of the global p-value in (3.18), without being overly conservative. We also propose a simple graphical tool to assess these conditions in practical applications.

4.1 REVIEW ON MAIN RESULTS ON POINT PROCESSES OF UP-CROSSINGS AND EXCEEDANCES

First of all, we introduce the distinction between upcrossings and *exceedances* of the sequence $\{W(\theta_r)\}$. Recall that the upcrossings of c_R by $\{W(\theta_r)\}$ over Θ_R are events of the type $\{W(\theta_{r-1}) \leq c_R, W(\theta_r) > c_R\}$; whereas we say that an exceedance of c_R by $\{W(\theta_r)\}$ occurs at θ_r if $\{W(\theta_r) > c_R\}$. An illustration of the difference between

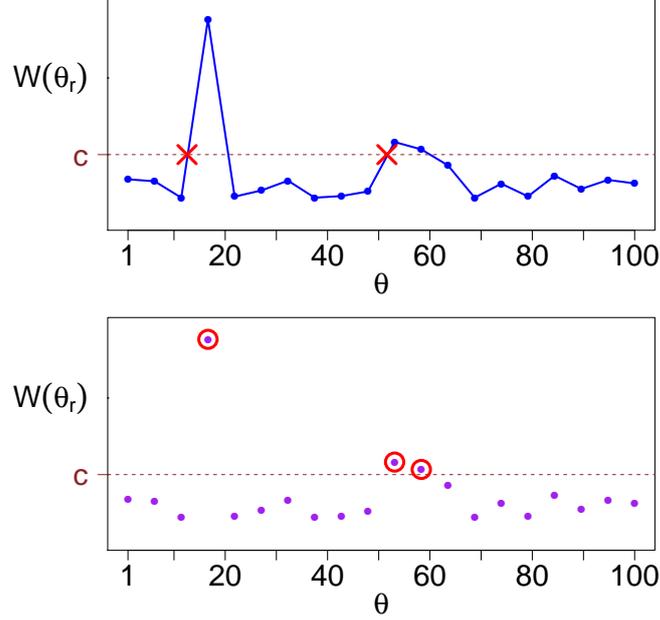


Figure 4.1: Upper panel: upcrossings (red crosses) of the threshold c by $\{W(\theta_r)\}$. Bottom panel: exceedances (red circles) of the threshold c by $\{W(\theta_r)\}$.

upcrossings and exceedances is given in Figure 4.1.

In order to characterize the behavior of $\{W(\theta_r)\}$, Conditions 3.1.1 and Condition 1.1.4, are replaced by the classical *mixing conditions* D, D^* and D' reported below, and further formalized in Hüsler [1983] and Falk et al. [2010, Ch. 9].

Condition 4.1.1 (Condition D). Let $F(\cdot)$ be the cumulative density function of each component of $\{W(\theta_r)\}$ and let $1 \leq r_1 < \dots < r_p < r'_1 < r'_2 < \dots < r'_q \leq R$ be any set of integers such that $r'_1 - r_p \geq d$, $J = \{r_j, j = 1, \dots, p\}$, $J' = \{r'_j, j = 1, \dots, q\}$, $B(J) = \{W(\theta_r) \leq c_R, r \in J\}$, $B(J') = \{W(\theta_{r'}) \leq c_R, r' \in J'\}$. Then we assume that

$$\sup_{J, J'} |P(B(J) \cap B(J')) - P(B(J))P(B(J'))| \leq v_{R,d} \quad (4.1)$$

where $v_{R,d_R^*} \rightarrow 0$ for some sequence d_R^* such that $d_R^*(1 - F(c_R)) \rightarrow 0$ as $R \rightarrow \infty$.

Condition 4.1.2 (Condition D^*). Let J^* be an ordered subset $\{r_1 \leq r \leq r_2\}$ of $\{1, \dots, R\}$ such that

$$\sum_{r \in J^*} P(W(\theta_r) > c_R) \leq \sum_{r \leq R} \frac{P(W(\theta_r) > c_R)}{R'}$$

for any integer R' . Let

$$\max_J \min_{J^* \subset J} \sum_{r < r' < r'+1 \in J^*} P(W(\theta_r) > c_R, W(\theta_{r'}) \leq c_R, W(\theta_{r'+1}) > c_R) \leq v_{R,R'}^* \quad (4.2)$$

and

$$\lim_{R' \rightarrow \infty} \limsup_{R \rightarrow \infty} R' v_{R,R'}^* = 0 \quad (4.3)$$

with $J^* \subset J$, such that $\sum_{r \in J \setminus J^*} P(W(\theta_r) > c_R) \leq t(R')/R'$, for all functions such that $t(R') \rightarrow 0$ as $R' \rightarrow \infty$.

Condition 4.1.3 (Condition D'). Let J be a ordered subset $\{r_1 \leq r \leq r_2\}$ of $\{1, \dots, R\}$ such that

$$\sum_{r \in J} P(W(\theta_r) > c_R) \leq \sum_{r \leq R} \frac{P(W(\theta_r) > c_R)}{R'},$$

for any integer R' . Let

$$\max_J \min_{J^* \subset J} \sum_{r < r' \in J^*} P(W(\theta_r) > c_R, W(\theta_{r'}) > c_R) \leq v_{R,R'}^* \quad (4.4)$$

and

$$\lim_{R' \rightarrow \infty} \limsup_{R \rightarrow \infty} R' v_{R,R'}^* = 0 \quad (4.5)$$

with $J^* \subset J$, such that $\sum_{r \in J \setminus J^*} P(W(\theta_r) > c_R) \leq t(R')/R'$, for all functions such that $t(R') \rightarrow 0$ as $R' \rightarrow \infty$.

Heuristically, Condition D , also known as *long-range independence mixing* condition, implies that independence is achieved for distant components of $\{W(\theta_r)\}$. Condition D^* , also known as *local independence*, prevents $\{W(\theta_r)\}$ from oscillating rapidly around the threshold c_R and thus does not allow the upcrossings of c_R to cluster over small intervals. Finally, Condition D' bounds the probability of multiple exceedances of c_R over small intervals. Notice that D' is simply a stronger version of Condition D^* and indeed, D' implies D^* . The long-range and local independence conditions are crucial to guarantee the Poisson nature of both the of upcrossing and exceedance processes of $\{W(\theta_r)\}$.

We denote with \dot{N}_{c_R} and \tilde{N}_{c_R} the process of exceedances and the process of upcrossings of c_R by $\{W(\theta_r)\}$, respectively. We require the following conditions to hold.

Conditions 4.1.4. We assume that on $[\mathfrak{L}; \mathfrak{U}]$, as $R \rightarrow \infty$ and $c_R \rightarrow \infty$

$$E[\dot{N}_{c_R}] = \sum_{r=1}^R P(W(\theta_r) \geq c_R) \rightarrow \dot{\mu} \leq \infty \quad (4.6)$$

$$E[\tilde{N}_{c_R}] = \sum_{r=1}^{R-1} P(W(\theta_r) \leq c_R, W(\theta_{r+1}) > c_R) \rightarrow \tilde{\mu} \geq 0 \quad (4.7)$$

where clearly, $0 \leq \tilde{\mu} \leq \dot{\mu} \leq \infty$.

Results 4.1.5 and 4.1.6 can be found in Hüsler [1983, 1986] and Falk et al. [2010, Ch. 9].

Result 4.1.5. Under Conditions 4.1.4, D and D^* , both \dot{N}_{c_R} and \tilde{N}_{c_R} converge in distribution to Poisson processes with means $\dot{\mu}$ and $\tilde{\mu}$, respectively, on $[\mathfrak{L}; \mathfrak{U}]$, as $R \rightarrow \infty$ and $c_R \rightarrow \infty$.

Result 4.1.6. Under Conditions 4.1.4, D and D' , both \dot{N}_{c_R} and \tilde{N}_{c_R} converge in distribution to the same Poisson process with mean $\dot{\mu} = \tilde{\mu}$ on $[\mathfrak{L}; \mathfrak{U}]$, as $R \rightarrow \infty$ and $c_R \rightarrow \infty$.

4.2 TOHM VIA BONFERRONI'S CORRECTION

We now turn to the practical implications of Results 4.1.5 and 4.1.6 reviewed in Section 4.1. First, we rewrite the expected number of exceedances in (4.6) in a form more familiar to practitioners

$$\begin{aligned} E[\dot{N}_{c_R}] &= \sum_{r=1}^R P(W(\theta_r) \geq c_R) = \sum_{r=1}^R P\left(W(\theta_r) \geq \max_{\theta_r \in \Theta_R} \{w(\theta_r)\}\right) \\ &= R \min_{\theta_r \in \Theta_R} P(W(\mathfrak{L}) \geq w(\theta_r)) \end{aligned} \quad (4.8)$$

The second line of (4.8) follows from the identical distribution of each component of the sequence $\{W(\theta_r)\}$, i.e., Condition (iii) among Conditions 3.1.1. Let $W(\mathfrak{L})$ be a random variable whose distribution is, exactly or asymptotically, the same of a test statistics $W_n(\theta_r)$ for each $\theta_r \in \Theta_R$. Then, the last line of (4.8) corresponds to the Bonferroni's correction introduced in (1.35), and more generally we write

$$p_{BF} = R \min_{\theta_r \in \Theta_R} P(W(\mathfrak{L}) \geq w(\theta_r)) \quad (4.9)$$

This implies that, since Bonferroni's correction corresponds to the average number of exceedances of $\{W(\theta_r)\}$, and the second term in the right hand side of (3.18) gives the expected number of upcrossings of $\{W(\theta_r)\}$, we expect (3.18) and (4.9) to be asymptotically equivalent when $E[\dot{N}_{c_R}]$ and $E[\tilde{N}_{c_R}]$ converge to the same limit. This can be shown by combining Results 4.1.5 and 4.1.6 with equations (4.6), (4.7) and (4.8).

Specifically, we have that, under Conditions D and D' , and as $R \rightarrow \infty$ and $c_R \rightarrow \infty$,

$$\text{by (4.7)} \quad E[\tilde{N}_{c_R}] \rightarrow \tilde{\mu} \quad (4.10)$$

$$\text{by (4.6) and (4.8)} \quad E[\dot{N}_{c_R}] = p_{BF} \rightarrow \dot{\mu} \quad (4.11)$$

$$\text{by Result 4.1.6} \quad \tilde{\mu} = \dot{\mu}. \quad (4.12)$$

These observations establish the following relationship between the global p-value and p_{BF} .

Theorem 4.2.1. *Let c_R and p_{BF} be defined as in (3.15) and (4.9), respectively, then*

$$P\left(\sup_{\theta_r \in \Theta_R} \{W(\theta_r)\} > c_R\right) \leq P(W(\mathfrak{L}) > c_R) + p_{BF}. \quad (4.13)$$

Under Condition 4.1.4, D and D' , (4.13) is asymptotically sharp, as $c_R \rightarrow \infty$. Further, if Condition 3.1.4 hold, the global p-value $P(\sup_{\theta \in \Theta} \{W(\theta)\} > c)$ is well approximated by p_{BF} .

Proof. Consider the sequence $\{W(\theta_r)\}$, and its process of upcrossings \tilde{N}_{c_R} . Because $E[\tilde{N}_{c_R}] \leq E[\dot{N}_{c_R}] = p_{BF}$ we can write

$$P\left(\sup_{\theta_r \in \Theta_R} \{W(\theta_r)\} > c_R\right) \leq P(W(\mathfrak{L}) > c_R) + E[\tilde{N}_{c_R}] \quad (4.14)$$

$$\leq P(W(\mathfrak{L}) > c_R) + p_{BF}, \quad (4.15)$$

where (4.14) follows from the same argument used to obtain (3.3) in Section 3.1. By Result 4.1.5, the Poisson nature of \tilde{N}_{c_R} guarantees that the right hand side of (4.14) is asymptotically equivalent to the left hand side as $c_R \rightarrow \infty$. By (4.12), D and D' , the right hand side of (4.14) and (4.15) are asymptotically equivalent, as $R \rightarrow \infty$ and $c_R \rightarrow \infty$. Thus (4.15) is also a sharp bound for the left hand side of (4.14). It follows that, if Condition 3.1.4 hold, both the right hand side of (4.14) and (4.15) are good approximations for $P(\sup_{\theta \in \Theta} \{W(\theta)\} > c)$ for large c (and c_R).

By the identical distribution of the components of $\{W(\theta_r)\}$ and following (4.8), we have

$$P(W(\mathfrak{L}) > c_R) = \min_{\theta_r \in \Theta_R} P(W(\mathfrak{L}) \geq w(\theta_r)) = \frac{p_{BF}}{R} \quad (4.16)$$

Substituting (4.16) into (4.13), we obtain

$$P\left(\sup_{\theta \in \Theta} \{W(\theta)\} > c\right) \approx \frac{R+1}{R} p_{BF} \approx p_{BF} \quad \text{as } R \rightarrow \infty. \quad (4.17)$$

□

It would be informative to quantify the rate at which $|p_{BF} - E[\tilde{N}_{c_R}]| \rightarrow 0$, and for which (4.14) and (4.15) are exchangeable. In Section A.2 of Appendix A we discuss a way to do this by considering the variational distance between the process of exceedances \tilde{N}_{c_R} and the Poisson process with mean $E[\tilde{N}_{c_R}]$ on $[\mathfrak{L}; \mathfrak{U}]$, under the conditions of Theorem 4.2.1.

4.2.1 ASSESSING D AND D' FOR GAUSSIAN AND RELATED SEQUENCES

In this section, we focus our attention to the Gaussian, χ_s^2 and $\bar{\chi}_{01}^2$ cases and studied in more details for Examples 2.1, 3.1 and 3.2 in Section 4.3. Similar results can be obtained for other Gaussian related stochastic processes.

Owing to their technical nature, it may appear that there is little hope for assessing the appropriateness of D , D^* and D' in practice. However, Hüsler [1983] and Raab [1997, Section 2.3.2] show, for Gaussian and χ_s^2 sequences, respectively, that if the so-called *Berman's condition* [Berman, 1964] in (4.18),

$$\sup_{|\theta_r - \theta_{r'}| > \tau} |\rho(\theta_r, \theta_{r'})| \log(\tau) \rightarrow 0 \quad \text{as } \tau \rightarrow +\infty. \quad (4.18)$$

holds, then D and D' also hold. Here, $\rho(\theta_r, \theta_{r'})$ represents the covariance function of the zero-mean and unit variance Gaussian sequence $\{Z(\theta_r)\}$ and, for $\{T(\theta_r)\}$, represents the covariance function of the underlying zero-mean and unit variance Gaussian sequences $\{Z_1(\theta_r)\}, \dots, \{Z_s(\theta_r)\}$, such that $\{T(\theta_r)\} = \sum_{j=1}^s \{Z_j(\theta_r)\}^2$.

Using (4.18), we can establish the following corollaries of Theorem 4.2.1

Corollary 4.2.2. *If (4.18) holds for the covariance function of $\{Z(\theta_r)\}$, then the results of Theorem 4.2.1 holds on $\{Z(\theta)\}$, as $R \rightarrow \infty$.*

Corollary 4.2.3. *If (4.18) holds for the covariance function of each element of the Gaussian sequences $\{Z_1(\theta_r)\}, \dots, \{Z_s(\theta_r)\}$ such that $\{T(\theta_r)\} = \sum_{j=1}^s \{Z_j(\theta_r)\}^2$, then the results of Theorem 4.2.1 holds on $\{T(\theta)\}$, as $R \rightarrow \infty$.*

Analogously, Corollary 4.2.4 follows for the $\bar{\chi}_{01}^2$ case. Suppose $\{K(\theta_r)\}$ can be rewritten as $\{K(\theta_r)\} = \{Z(\theta_r)\}^2 \mathbb{1}_{\{Z(\theta_r)\} \geq 0}$, for some $\{Z(\theta_r)\}$.

Corollary 4.2.4. *If (4.18) holds for the covariance function of the Gaussian sequences $\{Z(\theta_r)\}$ such that $\{K(\theta_r)\} = \{Z(\theta_r)\}^2 \mathbb{1}_{\{Z(\theta_r)\} \geq 0}$, then the results of Theorem 4.2.1 holds on $\{K(\theta)\}$, as $R \rightarrow \infty$.*

For Corollaries 4.2.2, 4.2.3 and 4.2.4 to be useful in practice, we must consider the appropriate Gaussian sequence for the evaluation of (4.18). Consider, for example, the LRT-process, $\{T_n(\theta)\}$, and the associated sequence of sub-test statistics $\{T_n(\theta_r)\}$ for the test in (1.2). For simplicity we let $\boldsymbol{\eta} = \eta$ be one-dimensional^a, i.e., $s = 1$.

The χ_1^2 or $\bar{\chi}_{01}^2$ asymptotic behavior of $\{T_n(\theta_r)\}$ under H_0 can be derived from the normalized score sequence

$$\{S_n^*(\eta_0, \hat{\phi}_0, \theta_r)\} = \left\{ \frac{\frac{\partial}{\partial \eta} \log \mathcal{L}_n(\eta_0, \hat{\phi}_0, \theta_r)}{\sqrt{I_n^*(\eta_0, \hat{\phi}_0, \theta_r)}} \right\} \quad (4.19)$$

where $\mathcal{L}_n(\eta_0, \hat{\phi}_0, \theta_r)$ is the profile-likelihood of $h(y, \eta, \phi, \theta_r)$, evaluated at $\hat{\phi}_0$, i.e., the MLE of ϕ under H_0 , η_0 is the value of η under H_0 and θ is fixed at θ_r . Whereas $I_n^*(\eta_0, \hat{\phi}_0, \theta_r) = I_{11} - I_{12}I_{22}^{-1}I_{21}$, with

$$I_n(\eta_0, \hat{\phi}_0, \theta_r) = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix} \quad (4.20)$$

being the Fisher information evaluated at η_0 and $\hat{\phi}_0$, with θ fixed at θ_r .

From Davies [1977] it follows that, under suitable conditions, the normalized score sequence $\{S_n^*(\eta_0, \hat{\phi}_0, \theta_r)\}$ with covariance function $\rho_n^*(\theta_r, \theta_{r'})$ is such that, as $n \rightarrow \infty$, $\{S_n^*(\eta_0, \hat{\phi}_0, \theta_r)\} \xrightarrow{d} \{S_{\eta_0}^*(\theta_r)\}$. Where $\{S_{\eta_0}^*(\theta_r)\}$ is a zero-mean unit variance Gaussian sequence with covariance function $\rho^*(\theta_r, \theta_{r'}) = \lim_{n \rightarrow \infty} \rho_n^*(\theta_r, \theta_{r'})$. Whether $\{T_n(\theta_r)\}$

^aHere, we restrict η to be univariate because applying Corollary 4.2.3 to multivariate $\boldsymbol{\eta}$ requires each of the components of the score vector $S(\boldsymbol{\eta}_0, \hat{\phi}_0, \theta_r)$ to be independent and the associated normalized sequences to have the same covariance function [Lindgren, 1980]. Unfortunately, this restrictive requirement limits the applicability of Corollary 4.2.3 when $\boldsymbol{\eta}$ is multivariate.

behaves asymptotically as $\{T(\theta)\}$ or $\{K(\theta)\}$ can be derived from $\{S_n^*(\eta_0, \hat{\phi}_0, \theta)\}^b$, and as in Corollaries 4.2.3 and 4.2.4, if (4.18) holds for $\rho_n^*(\theta_r, \theta_{r'})$ the result of Theorem 4.2.1 holds, as $n \rightarrow \infty$.

Notice that, when considering the signed-root-LRT process, $\{Q_n(\theta)\}$, given its asymptotic Gaussian nature, we can directly assess (4.18) on the associated sequence $\{Q_n(\theta_r)\}$, for large samples. Alternatively, we can refer to $\{S_{\eta_0}^*(\theta_r)\}$ also in this case. In applied settings, the assessment of (4.18), can be conducted with a simple graphical tool which we discuss in Section 4.3 and implement for Examples 2.1, 3.1 and 3.2.

4.3 TOHM, BONFERRONI AND BERMAN'S CONDITION

Section 4.2 discusses the relationship between TOHM and MHT conducted via Bonferroni's correction. In this framework, in addition to stringent significance requirement and large resolutions R , a crucial role is played by Berman's condition in (4.18). Specifically, if verified, this condition implies approximate equivalence between (3.18) and p_{BF} .

In practical applications the search range $[\mathfrak{L}; \mathfrak{U}]$ is bounded, in contrast to the theoretical assumption that $R \rightarrow \infty$, the number of grid points R is fixed, and consequently, τ in (4.18) is bounded by the length of the search window. Thus, it is useful to assess the validity of (4.18) under these circumstances on a case-by-case basis. A simple way to do this is to evaluate (4.18) using the covariance function $\rho_n^*(\theta_r, \theta_{r'})$ of the normalized score sequence, $\{S_n^*(\eta_0, \hat{\phi}_0, \theta_r)\}$, introduced in Section 4.2.1, over the specified grid Θ_R , with R selected as in Section 3.2.2.

In the first row of Figure 4.2 we assess (4.18) graphically for Examples 2.1, 3.1 and 3.2, using equally spaced grid points over the search ranges $[\mathfrak{L}; \mathfrak{U}]$ specified in Section 3.2.1, with $R = 50$, $R = 100$ and $R = 50$ respectively. In Example 3.1, the limit in (4.18) approaches zero over the range $[1; 35]$, thus we expect TOHM and MHT via Bonferroni to provide similar results. Conversely, in Examples 2.1 and 3.2, (4.18) does not hold over the respective search ranges.

In the second row of Figure 4.2, we compare the TOHM estimated bound in (3.18) and Bonferroni's bound for the three examples. Because the signed-root LRT, $\{Q_n(\theta)\}$, is used in Example 3.2 rather than the LRT, smaller values of c corre-

^bAs described in Davison [2003, p. 142], when testing on the boundary of the parameter space $P(S_n^*(\eta_0, \hat{\phi}_0, \theta) \geq 0) = \frac{1}{2}$ and thus $T_n(\theta) \rightarrow K(\theta)$, for each θ fixed.

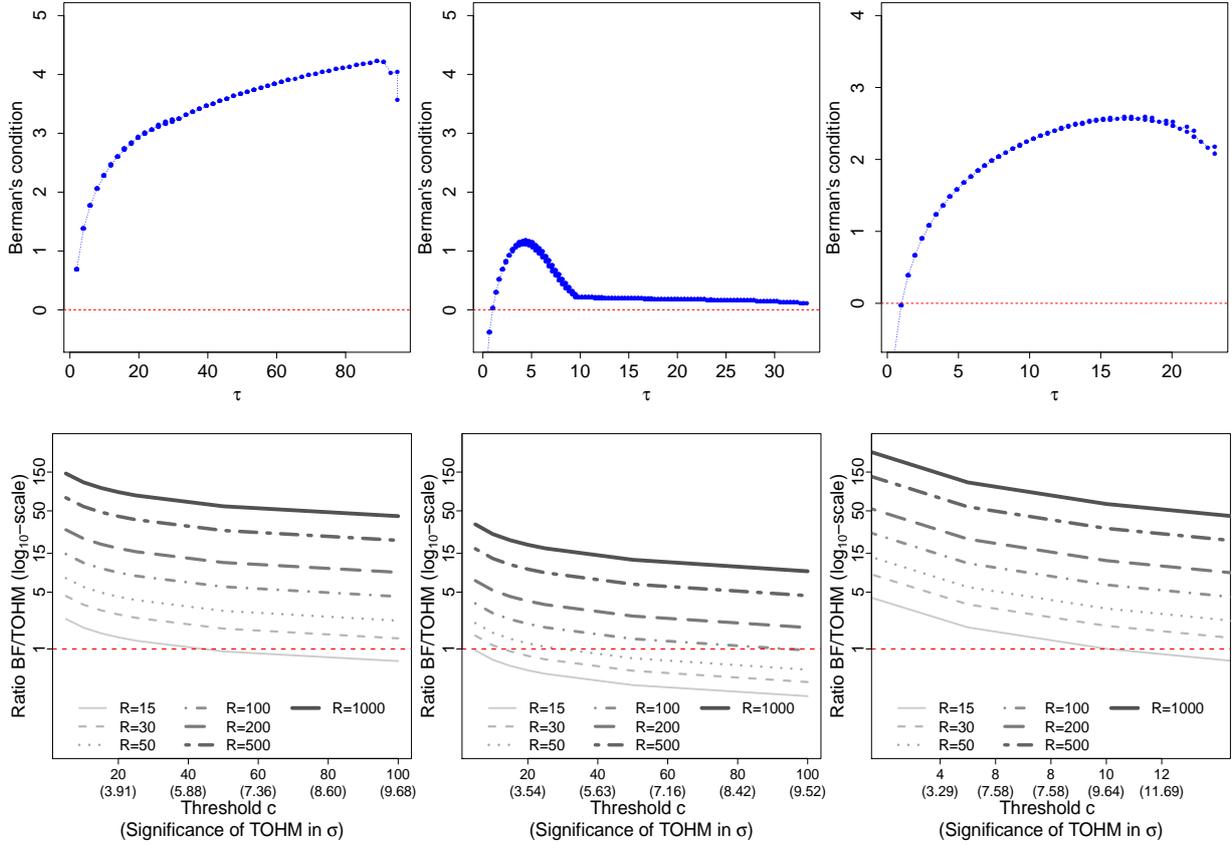


Figure 4.2: Top row: assessment of Berman's condition in (4.18) over the specified search ranges, with $R = 50, 100$ and 50 , for Example 2.1, 3.1 and 3.2, respectively. Bottom row: ratio of Bonferroni's bound and the bound in (3.18) at increasing values of c (and corresponding significance for TOHM), and considering different resolutions (grey curves). The first, second and third columns correspond to Example 2.1, 3.1 and 3.2, respectively.

spond to equally significant results. Thus, the statistical significance of TOHM is also reported (in σ) in the horizontal axes in the second row of Figure 4.2. Specifically, we plot the ratio of the two bounds for increasing values of c , using different grid sizes, R . As suggested by the Berman plots in the first row, for Example 3.1 (second column) the ratio of the bounds approaches one reasonably quickly, even for larger values of R . Interestingly, for lower resolutions Bonferroni is often less conservative than the TOHM bound. In Examples 2.1 and 3.2, however, Bonferroni is always more conservative than (3.18) when at least 30 tests are performed. All these plots suggest that the TOHM bound is preferable to Bonferroni with very high resolutions, i.e. $R \geq 500$.

Example	Test	Method	R	c_R	$\tilde{\theta}$	P-value (Significance)
Example 2.1	$H_0 : \eta = 0$	Bonferroni	50	21.021	27.265	$1.14 \cdot 10^{-4}$ (3.69σ)
	$H_1 : \eta > 0$	TOHM				$2.51 \cdot 10^{-5}$ (4.06σ)
	$H_0 : \eta = 1$	Bonferroni	50	0.606	27.890	> 1 (0.00σ)
	$H_1 : \eta < 1$	TOHM				$7.201 \cdot 10^{-1}$ (0.58σ)
Example 3.1	$H_0 : \eta = 0$	Bonferroni	100	38.326	3.404	$2.99 \cdot 10^{-8}$ (5.42σ)
	$H_1 : \eta > 0$	TOHM				$2.11 \cdot 10^{-8}$ (5.48σ)
Example 3.2	$H_0 : \xi = 0$	Bonferroni	50	11.826	31.266	$1.43 \cdot 10^{-30}$ (11.43σ)
	$H_1 : \xi \neq 0$	TOHM				$5.06 \cdot 10^{-31}$ (11.52σ)

Table 4.1: Summary of the results of TOHM and MHT via Bonferroni on real data for Examples 2.1, 3.1 and 3.2.

4.3.1 DATA ANALYSES

In this section we perform both TOHM and MHT via Bonferroni for Examples 2.1, 3.1 and 3.2. The results are summarized in Table 4.1, where evidence in favor of H_1 is reported in terms of p-value and respective σ -significance.

Not surprisingly, given the results of Section 4.3, for the dark matter search in Example 2.1 TOHM appears to be less conservative (4.06σ significance) than MHT via Bonferroni (3.69σ significance) in rejecting the hypothesis that the observed emission is due to a power-law distributed cosmic source. Because this example involves a non-nested models comparison, we invert the null of the hypotheses in order to avoid meaningless results (see Section 2.2 for more details). In the inverted test, the power-law model cannot be rejected with either TOHM (0.58σ) or MHT; given the low significance of the minimum of the local p-values (0.218 , result not shown), the Bonferroni’s bound for the p-value exceeds one (with $R = 100$). Both the fitted dark matter model and the fitted power-law cosmic source model are displayed in the central panel of Figure 3.1. In Example 2.1, the value of θ (i.e., the signal annihilation of the dark matter model) selected by both TOHM and Bonferroni is $\tilde{\theta} = 27.265\text{GeV}$. This is somewhat off from the true value used to simulate the data ($\theta = 35\text{GeV}$), perhaps because our analysis does not account for instrumental errors. Our analysis also only uses the spectral energy of the γ -ray

signals, whereas in practice the directions of the γ -ray would also be used, thus increasing the statistical power.

As expected, in the dark matter search problem of Example 3.1, we obtained a significance in favor of the presence of a dark matter emission of about 5.4σ using both TOHM and MHT. The signal location selected is close to the truth (3.5GeV), and the estimated model is plotted as a solid red line in the left panel of Figure 3.1; the signal location selected, $\tilde{\theta} = 3.404$, is indicated by the green dotted vertical line.

Finally, for the break-point regression model in Example 3.2, both TOHM and MHT give similar inferences (11.52σ and 11.43σ respectively) when rejecting the hypothesis of a linear model with no break-point. The similarity of the results is likely due to their extremely high statistical significance, i.e., we expect the bounds to coincide in the limit as $c \rightarrow \infty$. Comparing these results with those in the bottom right panel of Figure 4.2, even when performing 30 tests with 9.6σ significance ($c = 10$), Bonferroni is twice as conservative as the TOHM bound in (3.18). The fitted model is displayed in Figure 3.1 where the green triangle corresponds to the optimal break-point location, i.e., the maximum of the signed-root LRT process occurs at a mother's age of 31.266 years.

5

PRACTICAL GUIDELINES TO SELECT AMONG DIFFERENT INFERENTIAL PROCEDURES

In Chapter 4 we identify scenarios where, despite their difference in implementation, TOHM and MHT lead to the same inference. However, if Conditions 4.1.4 are not verified, (4.11) does not hold and thus Bonferroni's bound may easily diverge when the number of tests conducted, R , increases. Conversely, as discussed in Section 4.2.1, when R is relatively small, p_{BF} is often less conservative than the bound in (3.18) obtained by TOHM. Additionally, as alternative to the upcrossings heuristic of Chapter 3, the methods relying on volume-of-tube formulae, and introduced in Section 1.2, can be also be used to approximate excursion probabilities. Given the panacea of methods available, several factors must be taken in account in order to select the most convenient tool to be used in practical applications. The goal of this chapter is to provide general guidelines to practitioners for navigating the tradeoffs between statistical and computational efficiency when selecting a statistical procedure for signal detection. An exhaustive summary of the methods under comparison and the selection procedure described in this chapter can be found in [Algeri et al. \[2016b\]](#).

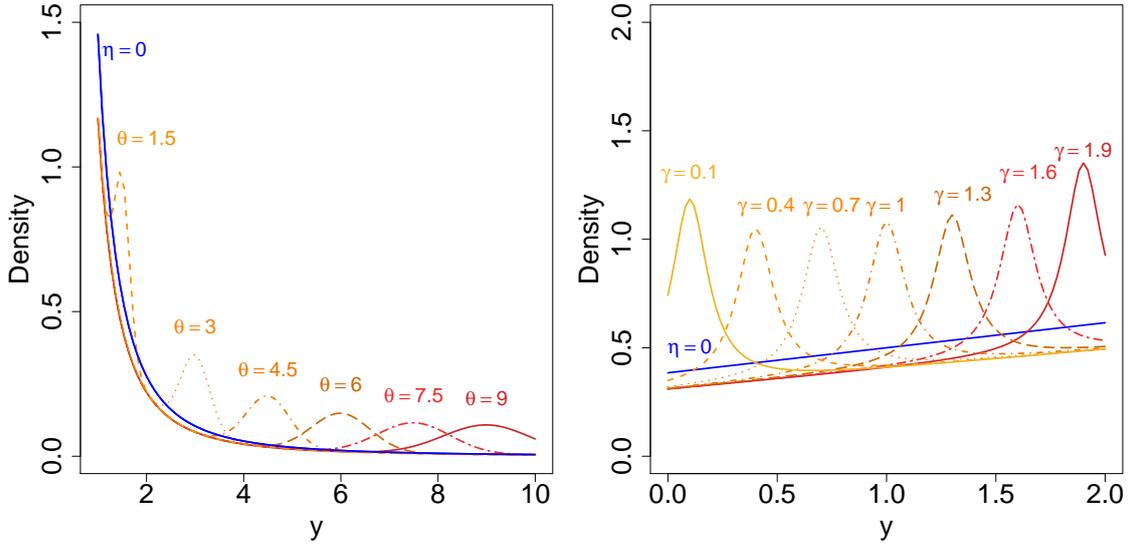


Figure 5.1: Left panel: probability density functions for Example 3.1 under H_0 (blue line) with $\phi = 1.4$ and H_1 (orange lines) with $\eta = 0.2$ and $\theta = 1.5, 3, 4.5, 6, 7.5, 9$. Right panel: probability density functions for Example 5.1 under H_0 (blue line) and H_1 (orange lines) with $\eta = 0.2$ and $\gamma = 0.1, 0.4, 0.7, 1, 1.3, 1.6, 1.9$.

5.1 CHOICE OF THE TEST STATISTICS AND STATISTICAL PROPERTIES

A fundamental result in probability theory states that the Score test and the LRT are asymptotically equivalent when the number of events, n , is large (i.e., for large sample sizes). As shown in Pilla et al. [2005], the same can be proven for the LRT and the normalized Score processes, $\{T_n(\theta)\}$ and $\{S_n^*(\theta)\}$ with components (1.4) and (1.17), respectively. Additionally, Takemura and Kuriki [2002] discuss asymptotic equivalence, for large c , of the approximations of excursion probabilities based on tube formulas, and those based on the expected number of upcrossings (or the expected Euler characteristic in multiple dimensions, see Chapter 6). As a consequence, we expect the methods of Gross and Vitells [2010] (hereafter GV) and Pilla et al. [2005], Pilla and Loader [2005] (hereafter PL) to be asymptotically equivalent for large n and large c .

However, differences may arise for small sample sizes and for only moderately large values of c ; in order to investigate this scenario, we consider two examples. As first example we refer to Example 3.1 introduced in Chapter 3, where the search involves a Gaussian signal on top of a power-law (Pareto type I) distributed background with unknown parameter ϕ . In our simulation studies, in order to ease the computation

involved in (1.18), we consider $y \in [1, 10]$, hence, the search region is also restricted to $\Theta \in [1, 10]$. As second example, we consider Example 5.1 introduced below.

Example 5.1. We consider the toy model in Pilla et al. [2005] where a Breit-Wigner emission, i.e., a Cauchy distributed signal, is superimposed on a linear background. The full model is

$$(1 - \eta) \frac{1 + 0.3y}{2.6} + \eta \frac{0.1}{k_\gamma \pi (0.01 + (y - \gamma)^2)} \quad (5.1)$$

where k_γ is a normalizing constant, $y \in [0; 2]$ and $\gamma \in (0; 2]$.

The probability density functions used in Examples 3.1 and 5.1 are plotted in Figure 5.1 considering different values of θ and γ . For both examples, we evaluate the probability of type I error (false detection rate), and the power for different values of the nuisance parameters θ and γ , and different sample sizes. Specifically, we let n to be 10, 50, 100, 200 and 500. The false detection rate and the power are obtained via Monte Carlo simulations from the null model ($\eta = 0$) and from the alternative model with $\eta = 0.2$, respectively. Since ϕ is unknown in Example 3.1, it can be estimated with the MLE $\hat{\phi}$ under H_0 . The simulations are then drawn from (3.20) with $\phi = \hat{\phi}$ via parametric bootstrap [Efron and Tibshirani, 1994]. In principle, the observed sample used to compute $\hat{\phi}$ could either come from the null or from the alternative model. Thus, in order to evaluate the consistency of (1.29) and (2.4) in both situations, two further sub-cases are needed. In Example 3.1a, we draw the “observed” sample from (3.20) with $\eta = 0$ and $\phi = 1.4$, i.e., in absence of new physics. In Example 3.1b, we draw the “observed” sample with $\eta = 0.2$, $\phi = 1.4$ and $\theta = 9$, i.e., when a signal is present at energy level 9GeV.

Results of the simulation studies appear in Figure 5.2. The columns of Figure 5.2 correspond to Example 3.1a, Example 3.1b and Example 5.1, respectively. In the first row, we report the simulated detection rates. The simulated test statistics $\sup_{\theta \in \Theta} \{T_n(\theta)\}$ and $\sup_{\theta \in \Theta} \{S_n^*(\theta)\}$ of GV and PL, respectively, were calculated for each of the 100,000 datasets generated from the null model. The values obtained were then compared to the nominal thresholds at 3σ , and calculated, as in (5.2) and (5.3), by setting the right hand side of (1.29) and (2.4) equal to $1 - \Phi(3) = 0.0013$ and solving for c , i.e.,

$$1 - \Phi(3) = \frac{\kappa_0}{2\pi} P(\chi_2^2 \geq c^2) + \frac{1}{2} P(\chi_1^2 \geq c^2) \quad (5.2)$$

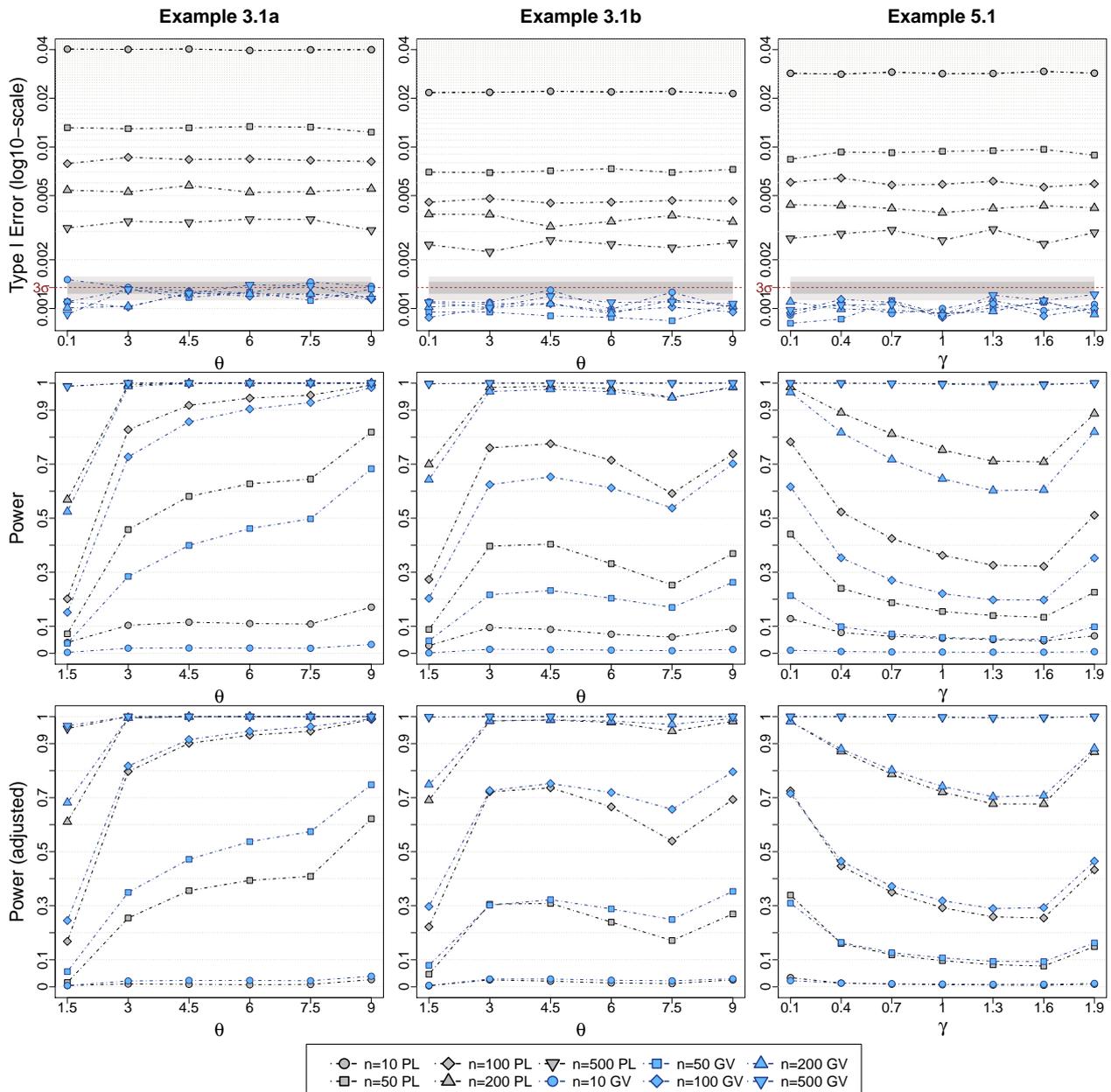


Figure 5.2: Simulated probability of type I error (top row), power (middle row) and adjusted power (bottom row) for Example 3.1a (first column), Example 3.1b (second column) and Example 5.1 (third column) with different sample size n over 100,000 simulations. The gray symbols corresponds to PL and the blue symbols to GV. Shaded areas indicate regions expected to contain 68% (dark gray) and 95% (light gray) of the symbols if the nominal type I error of 0.0013 holds.

$$1 - \Phi(3) = \frac{P(\chi_1^2 > c)}{2} + E[N_{c_0}]e^{-\frac{c-c_0}{2}}. \quad (5.3)$$

with $c_0 = 0.1$ and $E[N_{c_0}]$ estimated via Monte Carlo simulations as described in Chapter 3.

In the second row of Figure 5.2, we plot the power as a function of θ and γ for Examples 3.1 and 5.1, respectively. The procedure is the same as for the simulated false detection rates except the 100,000 datasets were generated from the alternative models with $\eta = 0.2$ while varying θ and γ . In the third row, we evaluate an adjusted version of the power; the simulated distributions of $\sup_{\theta \in \Theta} \{T_n(\theta)\}$ and $\sup_{\theta \in \Theta} \{S_n^*(\theta)\}$ are the same as those used in the plots in the second row, but instead of comparing them with the nominal thresholds obtained by solving (5.2) and (5.3), we compared them with their empirical (bootstrap) thresholds. The empirical threshold correspond to 99.87 quantiles of the 100,000 simulated values of $\sup_{\theta \in \Theta} \{T_n(\theta)\}$ and $\sup_{\theta \in \Theta} \{S_n^*(\theta)\}$ generated under H_0 for the first row of Figure 5.2, i.e., the empirical distributions of the test statistic under H_0 .

Looking at the first row of Figure 5.2, the simulated false detection rates associated with GV are always consistent with the nominal 3σ error rate. This is not the case for PL. Although the false detection curves approach the desired value as the sample size increases, they are always higher than expected. Looking at the second row of Figure 5.2, on the other hand, the simulated power when using $\sup_{\theta \in \Theta} \{S_n^*(\theta)\}$ is always higher than when considering $\sup_{\theta \in \Theta} \{T_n(\theta)\}$ as test statistics, at least for the smaller samples sizes. As expected, the difference between the power functions decreases when the sample size increases, leading to two identical curves at 500 counts. These results are, however, not sufficient to determine if PL is more powerful than GV. In fact, Figure 5.2 suggests that the increased power of PL is artificial, as it is due to an increase of the probability of a type I error. Conversely, when considering GV, the false detection rate is equal to or smaller than expected, and its power function approaches that of PL as the sample size increases.

As specified in (1.29), the right hand side is a valid approximation for the left hand side when c is large. In Examples 3.1a, 3.1b and 5.1 the values for c solving (5.2) are 3.896, 3.939 and 3.937, respectively. This implies that the approximation error in (1.29), i.e., $o\left(\frac{e^{-\frac{c^2}{2}}}{c}\right)$, is of the order of 10^{-4} , and thus the high false detection rate associated to PL is unlikely to be due to an underestimation of the 3σ nominal thresholds. Instead, it indicates that even a sample size of 500 is not sufficiently large to guarantee (1.21). This, however, does not invalidate the utility of this approach

Method	$\tilde{\theta}$	$\eta_{\tilde{\theta}}$	P-value	Significance
Unadjusted local	3.404	0.045	$2.99 \cdot 10^{-10}$	6.191σ
Bonferroni	3.404	0.045	$2.99 \cdot 10^{-8}$	5.419σ
Gross & Vitells	3.404	0.045	$2.10 \cdot 10^{-8}$	5.482σ
Pilla et al.	3.404	0.045*	$1.72 \cdot 10^{-9}$	5.909σ

*Obtained afterwards via MLE by fixing the signal location at $\tilde{\theta} = 3.404$ (see text).

Table 5.1: Implementing multiple hypothesis testing, GV and PL on the Fermi LAT simulation. For the multiple hypothesis testing case, the smallest of $R = 100$ (undadjusted local) p-values, Bonferroni’s bound on the global p-value, along with the bounds/approximations in (2.4) and (1.29), are reported with their respective statistic and σ -significances.

for larger n as shown in Pilla et al. [2005], Pilla and Loader [2005].

A more detailed comparison of the detection power of the two methods can be done by correcting the false detection rate (as in the third row of Figure 5.2). Specifically, we can use the empirical detection threshold when evaluating their power, i.e., the quantile of the simulated distribution of $\sup_{\theta \in \Theta} \{S_n^*(\theta)\}$ and $\sup_{\theta \in \Theta} \{T_n^*(\theta)\}$ which guarantees a false detection rate of 0.0013 (3σ significance). In this case, for all three examples and for all signal locations (values of γ or θ) considered, GV is equally or more powerful than PL.

Comparing the power functions in the second and third rows of Figure 5.2 with the density functions in Figure 5.1, we see that in Example 3.1 only the location of the source emission seems to affect the power. In particular, detection appears to be more difficult in high background areas of the spectrum, and thus the strength of the signal is weaker with respect to the background sources. These issues are overcome if at least 500 counts are available; in this case both procedure exhibit maximum detection power regardless the location or dispersion of the signal. For Example 5.1, the detection power of the testing procedures is affected by both the specific location of the signal and its spread over the search region. The power is higher when the signal is narrowly dispersed and is located in a region with low background.

Few computational difficulties arose when implementing the two methods under comparison. For PL, the most problematic step is the calculation of the geometric constant κ_0 in (1.30). This involves the numerical computation of nested integrals and it can significantly slow down the testing procedure for complicated models. In

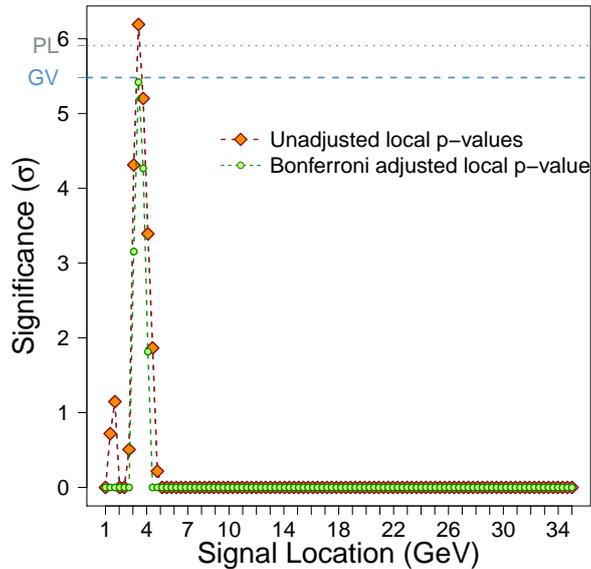


Figure 5.3: Unadjusted local p-values (orange diamonds), Bonferroni adjusted local p-values (green dots), global p-value (gray dotted line) implemented via PL and global p-value (blue dashed line) implemented via GV for the Fermi LAT simulation. The Bonferroni's bound is only slightly more conservative than the GV bound

the case of Examples 3.1 and 5.1, small search regions ($[1; 10]$ and $[0; 2]$ respectively) were chosen in order to speed up the computation of these integrals, which tended to diverge numerically over larger energy bands. The main difficulty with GV is associated with the maximization of the log-likelihood under H_1 . Here a multidimensional constrained optimization must be repeated for each point of the grid, Θ_R , for each Monte Carlo replicate used to estimate $E[N_{c_0}]$.

5.2 TOHM AND MHT ON FERMI LAT DATA

As a practical application, we perform PL, GV and multiple hypothesis testing via Bonferroni's correction considering Example 3.1 and the respective Fermi Large Area Telescope (LAT) simulation described in Section 3.2.1.

Results of the methods implemented are shown in Table 5.1 and Figure 5.3. In the multiple hypothesis testing analysis, the smallest of the local p-values is reported along with the respective estimates for the signal strength and location. As discussed in Chapter 4, the latter are approximately equivalent to those obtained via TOHM. When implementing PL, the test statistic $\sup_{\theta \in \Theta} \{S_n^*(\theta)\}$ is constructed under the

assumption that $\eta = 0$, and thus does not depend on the signal strength. However, it does depend on the location of the source emission, therefore the estimation of η under H_1 must be conducted once the signal location has been estimated (through MLE for instance). In our analysis, the location at which the processes $\{T_n(\theta)\}$ and $\{S_n^*(\theta)\}$ achieve their maximum, i.e., $\tilde{\theta}$ is the same at which the minimum among the local p-values is achieved, hence, the same follows for the MLE of the signal strength, namely $\hat{\eta}_{\tilde{\theta}}$.

When computing the local p-values the largest significance observed is 6.191σ , followed by PL with 5.909σ , GV with 5.482σ and finally the minimum of the Bonferroni adjusted local p-values leads to 5.419σ significance. Although PL gives to the most significant of the global p-values, it is difficult to interpret this result given the higher expected rate of false detections observed in the simulation study. The Bonferroni adjusted local p-value, over a set of $R = 100$ simultaneous tests, is only slightly more conservative than GV. The disparity between the two is expected to grow, however, as the number of grid points over the energy spectrum increases.

5.3 A SEQUENTIAL APPROACH TO SELECT AMONG DIFFERENT PROCEDURES

Denote with p_{PL} and p_{GV} the global p-values provided by the right hand side of (1.29) and (2.4) respectively. The PL and GV methods are typically used to overcome the over-conservativeness of the Bonferroni's bound, thus, one might expect p_{PL} and p_{GV} to be smaller or equal to p_{BF} . However, as shown in Theorem 6.2.1, this is not necessarily the case when the number R of tests conducted is small. Specifically, for large n and c ,

$$p_{\text{PL}} \approx p_{\text{GV}} \leq \frac{R+1}{R} p_{\text{BF}}. \quad (5.4)$$

On the basis of (5.4), and the results of the previous sections, it is possible to establish general guidelines for selecting the appropriate statistical testing procedure. The goal is to adhere to the prescribed false-positive rate as closely as possible while minimizing the computational effort. This can be accomplished by combining the simplicity of multiple hypothesis testing with the robustness of global p-values in a multi-stage procedure. Specifically, Figure 5.4 summarizes a simple sequence of steps where multiple hypothesis testing methods are implemented first, and the more time-consuming GV and PL are implemented only if simpler methods exhibit poor

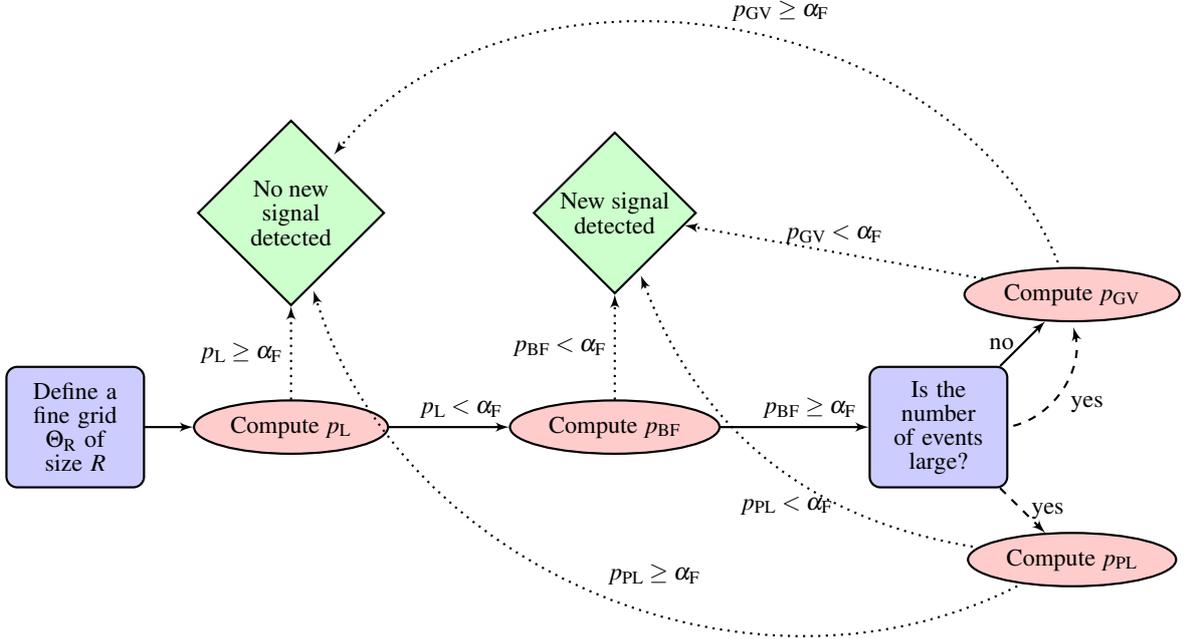


Figure 5.4: General guidelines for statistical signal detections via MHT and TOHM. Θ_R is the grid of possible signal-search locations. p_L is the minimum of the local p-values defined as in (1.31) and p_{BF} its Bonferroni adjusted counterpart in (4.9). α_F is the family-wise type I error rate. p_{PL} and p_{GV} are the global p-values provided by the right hand side of (1.29) and (2.4) respectively. Dashed arrows indicate that two actions are equally valid, and dotted lines lead to the final conclusion in terms of evidence in favor of the new signal.

type I error rates and/or power.

In order to implement the sequential approach, we first calculate the R unadjusted local p-values over the grid Θ_R ; the minimum of these p-values is denoted by p_L . Let α_F be the family-wise probability of type I error introduced in Section 1.3. If we observe $p_L > \alpha_F$ we fail to reject H_0 with any of the procedures and we can immediately conclude that we cannot reject H_0 . On the other hand, if $p_L \leq \alpha_F$, a correction for the simultaneous R tests is needed, and because of its easy implementation, we compute p_{BF} . Whereas, if $p_{BF} < \alpha_F$, then all methods reject H_0 , and we can claim evidence in favor of the new source. Conversely, if $p_{BF} \geq \alpha_F$ we should implement a method that is typically less conservative than Bonferroni's correction, when dealing with large significances (e.g. $3\sigma, 4\sigma, 5\sigma$), such as GV or PL. Specifically, on the basis of the simulations in Section 5.1, GV appears to be preferable for small sample sizes, as it provides a false-positive rate less than or

Method	Type I error	Power
Unadjusted local	0.03033	0.89502
Bonferroni	0.00040	0.45211
Gross & Vitells	0.00089	0.53159
Sequential approach	0.00087	0.53161

Table 5.2: Probability of type I error and power of the testing methods and sequential approach implemented on 100,000 simulated datasets from Example 3.1.

equal to α_F . For large sample sizes, PL and GV are equivalent, and the decision between GV and PL depends on the details of the models compared. As discussed in Section 5.1, PL requires extensive numerical integration which can diverge for large search windows Θ , while GV requires a small number of Monte Carlo simulations which might become troublesome for complicated models. Finally, if $p_{GV} < \alpha_F$ (or $p_{PL} < \alpha_F$) we can claim evidence in support of the new emission, whereas if $p_{GV} \geq \alpha_G$ (or $p_{PL} \geq \alpha_F$) we cannot claim that a signal has been detected.

The sequential approach described in Figure 5.4 involves choosing a procedure after having observed the inferential results. Thus, one might be concerned about possible “flip-flopping” similar to that described in [Feldman and Cousins \[1998\]](#) in the context of confidence intervals^a, i.e., the selection among different inferential procedures on the basis of their results may artificially inflate the type I error rate and the power. As argued below, however, this is not the case for the approach illustrated in Figure 5.4.

By virtue of Theorem 6.2.1, both the type I error and the power of the sequential approach are approximately equivalent to those of GV (or PL) for large values of R . For simplicity, we hereafter suppose that the sample size is sufficiently large to guarantee asymptotic equality of GV and PL, and GV is used when $p_{BF} \geq \alpha_F$. (The results below follow in exactly the same way however, if PL is used instead of GV.) Let $\tilde{\alpha}$ be the overall false detection rate associated with the sequential approach proposed in Figure 5.4, and consider the events

$$BF_1 = \{\text{Reject } H_0 \text{ at level } \alpha_F \text{ with Bonferroni}\}$$

$$GV_1 = \{\text{Reject } H_0 \text{ at level } \alpha_F \text{ with GV}\}.$$

^aIn high energy physics researchers often decide to report upper limits instead of confidence intervals on the basis of the results of the analysis being conducted. [Feldman and Cousins \[1998\]](#) show that this approach leads to wrong coverage and propose a computational solution to build confidence which guarantees correct coverage.

	Unadjusted local	Bonferroni	Gross & Vitells
Bkg only Time (secs)	97056 0.974	37 0.000	2907 136.282
Bkg+sig Time (secs)	10496 1.061	45210 0.000	44294 137.532

Table 5.3: Summary on the analysis of 100,000 simulated datasets from Example 3.1. We report the number of times each testing method is used by the sequential approach to make a final decision at 3σ , and the respective average computational times. The first two lines refer to the background only simulations and whereas the last two lines correspond to the background + signal simulations.

Because the sequential approach rejects H_0 when either Bonferroni or GV does so, it follows that

$$\begin{aligned}
\tilde{\alpha} &= P(\text{BF}_1 \cup \text{GV}_1 | H_0) \\
&= P(\text{BF}_1 | H_0) + P(\text{GV}_1 | H_0) - P(\text{BF}_1 \cap \text{GV}_1 | H_0) \\
&= P(\text{BF}_1 | H_0) + P(\text{GV}_1 | H_0) - P(\text{GV}_1 | \text{BF}_1, H_0)P(\text{BF}_1 | H_0).
\end{aligned}$$

By (5.4), if H_0 is rejected by Bonferroni then, for large R , it is also rejected by GV and thus,

$$P(\text{GV}_1 | \text{BF}_1, H_0) \approx 1,$$

from which it follows that $\tilde{\alpha} \approx P(\text{GV}_1 | H_0)$, where $P(\text{GV}_1 | H_0)$ is the probability of type I error of GV. Conversely, for small R , GV is more conservative than Bonferroni, and thus $P(\text{BF}_1 | \text{GV}_1, H_0) \approx 1$. Rewriting $P(\text{GV}_1 | \text{BF}_1, H_0)$ via Bayes theorem we obtain

$$P(\text{GV}_1 | \text{BF}_1, H_0) = \frac{P(\text{BF}_1 | \text{GV}_1, H_0)P(\text{GV}_1 | H_0)}{P(\text{BF}_1 | H_0)} \approx \frac{P(\text{GV}_1 | H_0)}{P(\text{BF}_1 | H_0)},$$

from which it follows that $\tilde{\alpha} \approx P(\text{BF}_1 | H_0)$.

To further investigate these aspects, we apply the sequential approach to a set of 100,000 simulated datasets from the model in Example 3.1 with ϕ fixed at 1.4. For each dataset we first simulate 2000 background only events and then we simulate 30 additional events from a Gaussian source centered at 9GeV. For both the 100,000 background only datasets and the 100,000 background plus source datasets we compute unadjusted local p-values, Bonferroni's corrections, and GV. The power of the

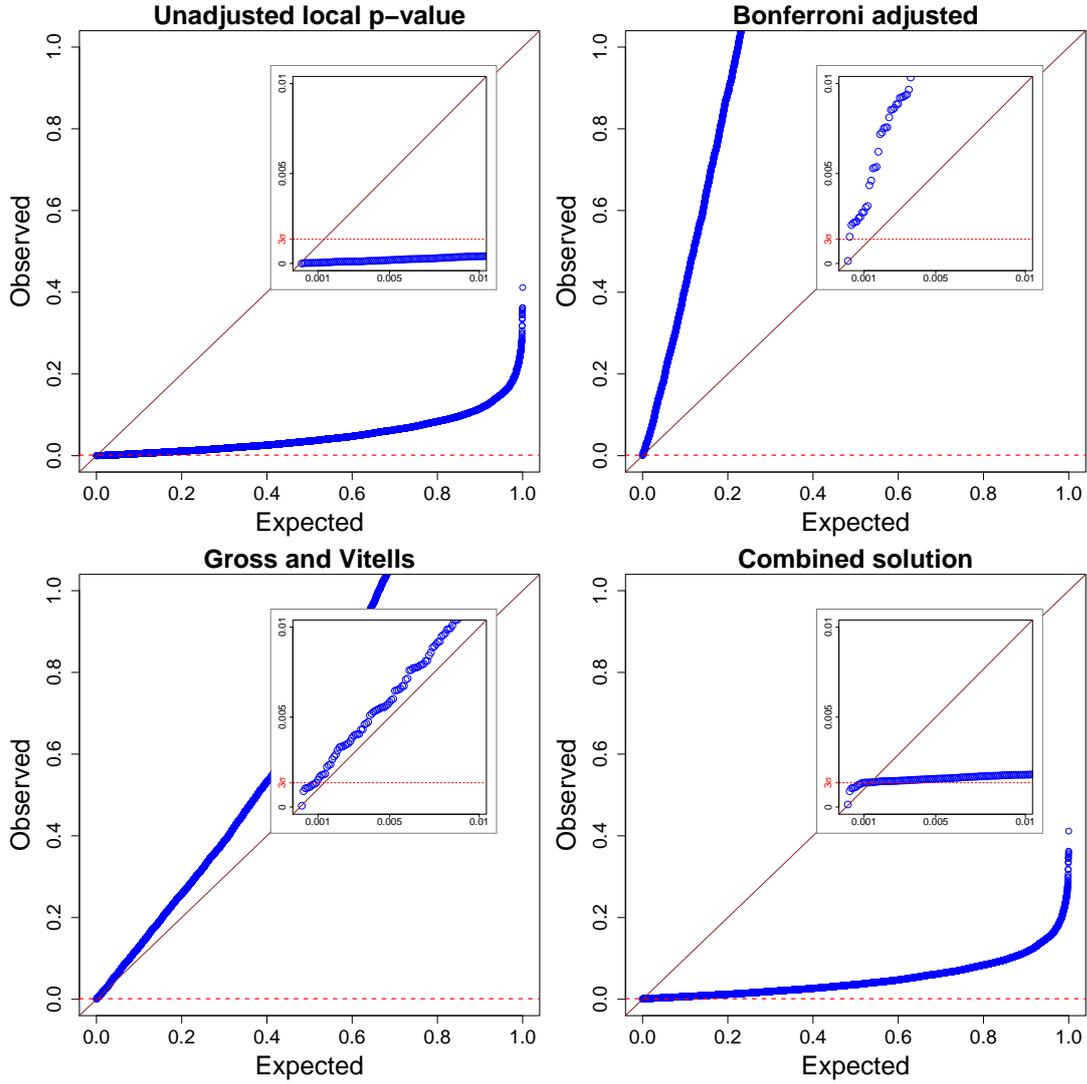


Figure 5.5: QQ-plots for the unadjusted local, Bonferroni’s bound and GV p-values computed for the 100,000 simulated background-only datasets from Example 3.1. Each dataset considers 2000 background only events. The p-values selected via the sequential procedure in Figure 5.4 are also reported. Each set of p-values is compared with the expected quantiles of a Uniform distribution on $[0, 1]$. The inlayed plots in each panel magnify the important range of the p-value distributions near zero.

sequential approach can be obtained in a similar manner by considering the events

$$L_1 = \{\text{Reject } H_0 \text{ at level } \alpha_F \text{ with local p-values}\}$$

$$GV_1 = \{\text{Reject } H_0 \text{ at level } \alpha_F \text{ with GV}\}.$$

Table 5.3 reports the number of times each of the testing procedures considered is selected by the sequential approach to make a final decision at the 3σ significance level. The average computational times for each method are also reported. In the

presence of source emission, the most computationally expensive method GV, and implemented via the R package `NONnest` [Algeri, 2015], was used only about 44% of the time, leading to a computational gain of about 89 days over the 100,000 simulations. Conversely, in absence of the signal, GV was used about 2.9% of the time, leading to a computational gain of about 155 days. In order to assess the robustness of the method with respect to the desired statistical properties, we computed the false discovery rate and the power using nominal levels at 3σ significance. The results are presented in Table 5.2. As discussed above, the sequential approach exhibits statistical properties which are approximately equivalent to those of GV (or PL). As expected, the small discrepancies between the two methods are due to the fact that in 0.375% of the replications $p_{GV} > p_{BF}$. When removing these cases from the analysis, both the probability of a Type I error and the power of the sequential approach coincide with those of GV.

Finally, Figure 5.5 displays the p-values computed with each procedure on each of the 100,000 simulated background-only datasets. Ideally a p-value will follow a uniform distribution on the unit interval under repeated sampling of data under H_0 ; this insures that the method will have the target Type I error rate. In the QQ-plots in Figure 5.5, the p-values will fall along the 45° line if they follow a uniform distribution. If they deviate above this line, the procedure is conservative and if they deviate below the procedure will exhibit too many false positives. As expected, the unadjusted local p-values are always smaller than their expected values assuming uniform distribution, whereas both Bonferroni and GV are conservative. The sequential approach leads to an intermediate situation in which the p-values are over-conservative up to the significance level α_F adopted at each step of the algorithm in Figure 5.4 (3σ in Figure 5.5), whereas the p-values become under-conservative above α_F , i.e., only for uninteresting cases.

6

TESTING ONE HYPOTHESIS MULTIPLE TIMES: THE MULTIDIMENSIONAL CASE

In Chapter 3 a simple expansion for the expected number of upcrossings of $\{W(\theta)\}$, is used to bound/approximate global p-values. In this chapter, we extend these results to the multidimensional setting, i.e., we let $\theta \in \Theta \subset \mathbb{R}^D$, for arbitrarily large D . Similarly to the one-dimensional case, we consider the supremum of a random field indexed by θ as test statistic, and instead of the number of upcrossings we use the *mean Euler characteristic (EC) of the excursion set* to approximate the resulting global p-value. In order to overcome the difficulties associated to the calculation of the mean EC, we discuss a simple algorithm to compute the EC in multiple dimensions. This leads to an highly generalizable computational tool to perform inference under non-standard regularity conditions [Algeri and van Dyk, 2018].

6.1 A QUICK REVIEW ON RANDOM FIELDS AND GEOMETRY

To generalize the framework of Chapter 1, let Y be a random variable with probability density $h(y, \boldsymbol{\eta}, \boldsymbol{\theta})$ and let $\mathbf{y} = (y_1, \dots, y_n)$ be a random sample, each component of which is distributed as Y . Suppose we wish to test a one-sided alternative hypothesis

$$H_0 : \boldsymbol{\eta} = \boldsymbol{\eta}_0 \quad \text{versus} \quad H_1 : \boldsymbol{\eta} > \boldsymbol{\eta}_0, \quad (6.1)$$

analogous results follow when testing H_0 versus $H_1 : \boldsymbol{\eta} \neq \boldsymbol{\eta}_0$ or $H_1 : \boldsymbol{\eta} < \boldsymbol{\eta}_0$.

We let $\boldsymbol{\eta} \in \Xi$ index the hypotheses in that $\boldsymbol{\eta} = \boldsymbol{\eta}_0$ defines H_0 , and let $\boldsymbol{\theta}$ be the nuisance parameter that is only defined under H_1 , it has no value under H_0 , and such that $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^D$, with $D \geq 1$. In Chapter 3, tests of hypothesis such as those in (1.15), (2.3) or (3.22) are performed by means of a stochastic process, namely $\{W(\boldsymbol{\theta})\}$, indexed by the one-dimensional parameter θ . In order to extend these results to the case where $\boldsymbol{\theta}$ is multidimensional, we let $\{W(\boldsymbol{\theta})\} = \{W(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$ be a D -dimensional random field^a indexed by $\boldsymbol{\theta}$, and in order to test (6.1) we consider the global p-value

$$P\left(\sup_{\boldsymbol{\theta} \in \Theta} \{W(\boldsymbol{\theta})\} > c\right), \quad c \in \mathbb{R}. \quad (6.2)$$

In the one-dimensional setting, $P(\sup_{\theta \in \Theta} \{W(\theta)\} > c)$ is modelled via the probability of having at least one upcrossing of c . The definition of upcrossings in Section 1.1.2, however, is unhelpful in the multidimensional setting. For smooth random fields for instance, if $\{W(\boldsymbol{\theta})\}$ crosses c at a point $\boldsymbol{\theta}_0$, it also crosses c at all the nearby points, leading to an infinite number of crossings. Therefore, our first aim is to identify a generalization of the number of upcrossings in the context of random fields.

Following Hasofer [1978], one possibility is to consider the number of local maxima^b of $\{W(\boldsymbol{\theta})\}$ that exceed c , namely M_c , hence

$$P\left(\sup_{\boldsymbol{\theta} \in \Theta} \{W(\boldsymbol{\theta})\} > c\right) = P(M_c \geq 1) \leq E(M_c). \quad (6.3)$$

Unfortunately, analytical expressions for $E(M_c)$ are known in literature only asymptotically in c , and thus cannot be exploited to derive multidimensional counterparts of Results 3.1.2 and 3.1.3, which would rely on evaluating $E(M_{c_0})$ at an arbitrarily small c_0 . A quantity that is more amenable and for which analytical expressions are known exactly, is the *expected Euler characteristic (EC) of the excursion set of $\{W(\boldsymbol{\theta})\}$ above c* . A clear description of the EC requires a few concepts from geometry that we now summarize [see Adler, 2000].

^aThe term “field” is typically used to indicate a stochastic process whose parameter space is multidimensional [Vanmarcke, 2010, p. 21].

^bWe are interested in scenarios where local maxima becomes rarer and rarer as $c \rightarrow +\infty$. Hence, we are implicitly assuming that no ridges above c occur. However, the procedure to be discussed aims to approximate the number of local maxima with the number of connected components above c , hence it also covers situations where, instead of isolated local maxima, sets of local maxima (ridges) are present.

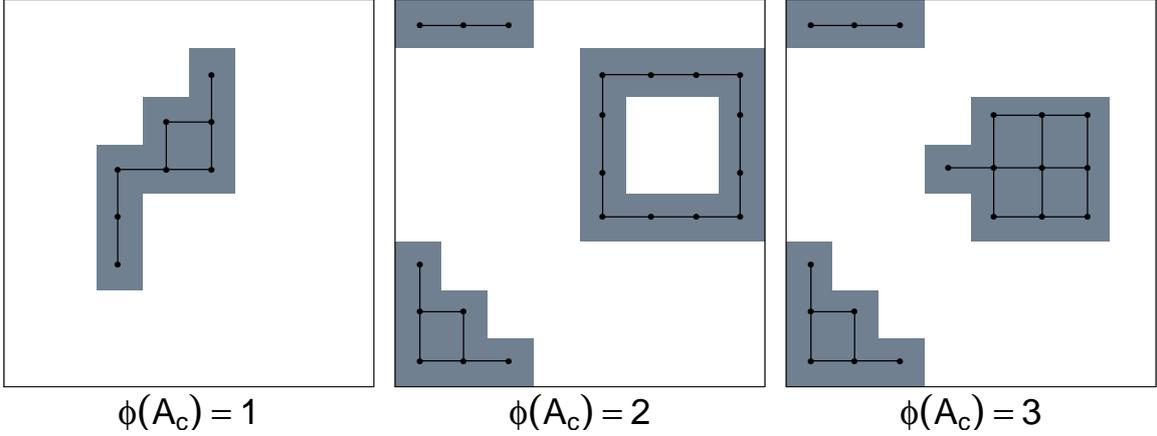


Figure 6.1: The shaded regions illustrate three possible excursion sets \mathcal{A}_c . The Euler characteristic (EC) of \mathcal{A}_c in the left, central and right panels are 1, 2 and 3, respectively. The EC can be obtained by counting the number of connected components less the number of holes of \mathcal{A}_c . Alternatively, considering a quadrilateral mesh of the image (black points and black edges), the same Euler characteristic is given by the number of points less the number of edges plus the number of squares.

Definition 6.1.1. *The excursion set of $\{W(\boldsymbol{\theta})\}$ above c is the set of points*

$$\mathcal{A}_c = \{\boldsymbol{\theta} \in \Theta : W(\boldsymbol{\theta}) \geq c\}. \quad (6.4)$$

Definition 6.1.2. *The Euler characteristic, $\phi(A)$, of a compact set $A \subset \mathbb{R}^D$ is the additive, integer-valued functional of A uniquely determined by the following properties:*

$$\phi(A) = \begin{cases} 1 & \text{if } A \text{ is homeomorphic to a } D\text{-dimensional sphere;} \\ 0 & \text{otherwise.} \end{cases} \quad (6.5)$$

and

$$\phi(A \cup B) = \phi(A) + \phi(B) - \phi(A \cap B).$$

Intuitively, in two dimensions the EC of \mathcal{A}_c is its number of connected components less its number of “holes”, see Figure 6.1. As noted by Hasofer [1978], the maxima of $\{W(\boldsymbol{\theta})\}$ above large values of c can be approximated by elliptic paraboloids, which correspond to connected components of \mathcal{A}_c . Hence, for large c , the EC approximately equals the number of connected components, and thus also approximately equals the number of local maxima above c . It follows from (6.3) that, as $c \rightarrow \infty$,

$$P\left(\sup_{\boldsymbol{\theta} \in \Theta} \{W(\boldsymbol{\theta})\} > c\right) \approx E[\phi(\mathcal{A}_c)]. \quad (6.6)$$

Gaussian case	$\rho_0(c) = 1 - \Phi(c) \qquad \rho_1(c) = \frac{e^{-c^2/2}}{2\pi}$ $\rho_2(c) = \frac{e^{-c^2/2}}{(2\pi)^{3/2}} \qquad \rho_3(c) = \frac{(c^2-1)e^{-c^2/2}}{(2\pi)^2}$ $\rho_4(c) = \frac{(c^3-3c)e^{-c^2/2}}{(2\pi)^{5/2}} \qquad \rho_5(c) = \frac{(c^4-4c^2+3)e^{-c^2/2}}{(2\pi)^3}$
χ_s^2 case	$\rho_0(c) = 1 - F_\chi(c) \qquad \rho_1(c) = \frac{c^{\frac{s-1}{2}}}{\Gamma(\frac{s}{2})} \sqrt{\frac{2}{\pi}} e^{-\frac{c}{2}}$ $\rho_2(c) = \left(\frac{c}{2}\right)^{\frac{s}{2}-1} \frac{e^{-\frac{c}{2}}}{2\pi} [c - (s-1)\mathbb{1}_{\{s \geq 2\}}]$ $\rho_3(c) = \frac{c^{\frac{s-3}{2}} e^{-\frac{c}{2}}}{(2\pi)^{3/2} \Gamma(\frac{s}{2}) 2^{\frac{s-2}{2}}} \left\{ (s-1)[(s-2)\mathbb{1}_{\{s \geq 3\}} - 2c\mathbb{1}_{\{s \geq 2\}}] + (c^2 - c)\mathbb{1}_{\{s \geq 1\}} \right\}$

Table 6.1: EC densities, $\rho_d(c)$, of Gaussian and χ^2 random fields. $\Phi(\cdot)$ is the cumulative function of a standard normal, $F_\chi(\cdot)$ is the cumulative function of a χ_s^2 and $\mathbb{1}\{\cdot\}$ is the indicator function. See Adler and Taylor [2009, p. 426] for higher order EC densities.

Worsley [1994, 1995] and Adler [2000], among others, give analytical expressions for $E[\phi(\mathcal{A}_c)]$, but they are often limited by regularity conditions on $\{W(\boldsymbol{\theta})\}$, \mathcal{A}_c and Θ , or by the dimension of Θ . A more generalizable approach is given by the seminal work of Taylor and Adler [2003], Adler and Taylor [2009] and Taylor and Worsley [2008]. They provide a convenient expansion of $E[\phi(\mathcal{A}_c)]$ for smooth Gaussian-related random fields on smooth manifolds with piecewise smooth boundaries. Specifically, we assume that $\{W(\boldsymbol{\theta})\}$ can be written as a function of i.i.d copies of a Gaussian random field $\{Z(\boldsymbol{\theta})\}$, with mean zero and variance one, and we require Conditions 6.1.3 to hold.

Conditions 6.1.3.

- (i) Θ is a compact manifold differentiable up to at least the third order;
- (ii) $\{Z(\boldsymbol{\theta})\}$ has almost surely uniformly continuous partial derivatives up to the second order, with finite second moments in an open neighbourhood of Θ ;
- (iii) the joint distribution of $\{Z(\boldsymbol{\theta})\}$ and its partial derivatives is non-degenerate.

see Taylor and Adler [2003, p. 547], for a more detailed formalization of Conditions 6.1.3 in geometric terms.

From Taylor and Adler [2003], it follows that, under Conditions 6.1.3,

$$E[\phi(\mathcal{A}_c)] = \sum_{d=0}^D \mathcal{L}_d(\Theta) \rho_d(c), \quad (6.7)$$

where the $\rho_d(c)$ are functionals known as EC densities, and only depend on the (identical) marginal distribution of each $W(\boldsymbol{\theta})$ in $\{W(\boldsymbol{\theta})\}$; for example, $\rho_0(c) = P(W(\boldsymbol{\theta}) > c)$. Closed-form expressions of $\rho_d(c)$ are available in literature for Gaussian, χ^2 , F and other Gaussian-related random fields [Taylor and Adler, 2003, Adler and Taylor, 2009, Taylor and Worsley, 2008]. The functionals $\mathcal{L}_d(\Theta)$ are known as the Lipschitz-Killing curvatures of Θ . Intuitively, they measure the intrinsic volume of Θ , i.e., they account for its volume, surface area, and boundaries. Their analytical forms typically rely on the covariance structure and partial derivatives of $\{W(\boldsymbol{\theta})\}$.

Unfortunately, obtaining closed-form expressions for $\mathcal{L}_d(\Theta)$ is challenging for non-isotropic fields [Adler and Taylor, 2009]. Even in the isotropic case this may require tedious calculations and knowledge of the distribution of the derivatives of $\{W(\boldsymbol{\theta})\}$. In the next two sections we introduce a novel approach to estimate the $\mathcal{L}_d(\Theta)$ in (6.7), and consequently, to compute the approximation for the global p-value in (6.6).

The error rate associated with (6.6) is exponentially small in the Gaussian case [Taylor et al., 2005], specifically,

$$P\left(\sup_{\boldsymbol{\theta} \in \Theta} \{Z(\boldsymbol{\theta})\} > c\right) = E[\phi^Z(\mathcal{A}_c)] + O(e^{-\frac{vc^2}{2}}) \quad (6.8)$$

for some $v > 1$ and with $E[\phi^Z(\mathcal{A}_c)]$ being the expected EC of the excursion set of $\{Z(\boldsymbol{\theta})\}$ with respect to c . However, no quantification of errors are available for non-Gaussian fields [Taylor and Worsley, 2008].

6.2 METHODOLOGICAL SETUP

In this section we extend the results of Section 3.1 with the goal of efficiently computing the right hand side of (6.7). This can be done following the approach implemented by Vitells and Gross [2011] in two dimensions, which we formalize and

extend to an arbitrary large dimension in Result 6.2.1 and Result 6.2.2.

Result 6.2.1. *Let $c \in \mathbb{R}$, and define a sequence of constants $c_1 \neq c_2 \neq \dots \neq c_D$, with $c_k \in \mathbb{R}$ for $k = 1, \dots, D$. If (6.7) hold, then,*

$$E[\phi(\mathcal{A}_c)] = \mathcal{L}_0(\Theta)P(W(\theta) > c) + \sum_{d=1}^D \mathcal{L}_d^*(\Theta)\rho_d(c), \quad (6.9)$$

where $\mathcal{L}_d^*(\Theta)$ are the solutions of the system of D linear equations

$$\begin{cases} E[\phi(\mathcal{A}_{c_1})] - \mathcal{L}_0(\Theta)\rho_0(c_1) & = \sum_{d=1}^D \mathcal{L}_d(\Theta)\rho_d(c_1) \\ E[\phi(\mathcal{A}_{c_2})] - \mathcal{L}_0(\Theta)\rho_0(c_2) & = \sum_{d=1}^D \mathcal{L}_d(\Theta)\rho_d(c_2) \\ & \vdots \\ E[\phi(\mathcal{A}_{c_D})] - \mathcal{L}_0(\Theta)\rho_0(c_D) & = \sum_{d=1}^D \mathcal{L}_d(\Theta)\rho_d(c_D), \end{cases} \quad (6.10)$$

with \mathcal{A}_{c_k} being the excursion sets of $\{W(\theta)\}$ above the constants c_k and $E[\phi(\mathcal{A}_{c_k})]$ their expected EC.

Proof. The proof is straightforward since the expansion for the expected EC in (6.7) holds for any value c . The Lipschitz-Killing curvature for $d = 0$, $\mathcal{L}_0(\Theta)$, corresponds to the Euler characteristic of Θ [Taylor and Worsley, 2008]. Since Θ is not random, $\mathcal{L}_0(\Theta)$ is known (e.g., $\mathcal{L}_0(\Theta)$ is 0, 1, 1 or 2 if Θ is a circle, a disc, a square or a cube, respectively) and does not need to be estimated. Finally, the EC densities $\rho_d(\cdot)$ are known for all $d = 0, \dots, D$. Thus, (6.10) can be seen as a system of D linear equations with D unknowns, i.e., $\mathcal{L}_1(\Theta), \dots, \mathcal{L}_D(\Theta)$, and its solution is given by the vector $\mathcal{L}_1^*(\Theta), \dots, \mathcal{L}_D^*(\Theta)$. \square

Result 6.2.2 follows from (6.6), (6.7) and (6.9).

Result 6.2.2. *Under the conditions of Result 6.2.1, (6.6), (6.7) and (6.9) together imply*

$$P\left(\sup_{\theta \in \Theta} \{W(\theta)\} > c\right) \approx \mathcal{L}_0(\Theta)P(W(\theta) > c) + \sum_{j=1}^D \mathcal{L}_j^*(\Theta)\rho_j(c) \quad (6.11)$$

as $c \rightarrow \infty$.

In order to compute the solution $\mathcal{L}_1^*(\Theta), \dots, \mathcal{L}_D^*(\Theta)$ of (6.10) we estimate each $E[\phi(\mathcal{A}_{c_k})]$ via a Monte Carlo simulation; details are given in Section 6.4. We could

also estimate $\mathcal{L}_d(\Theta)$, $d = 1, \dots, D$ via regression as described in Adler et al. [2017]. Specifically, in Adler et al. [2017] the Lipschitz-Killing curvatures $\mathcal{L}_d(\Theta)$ are treated as unknown regression coefficients, whereas the response vector and the design matrix of the regression model are given by $E[\phi(\mathcal{A}_c)]$ and the EC densities $\rho_d(c)$, respectively, evaluated over different realizations of the random field under study, and for different values of c . The latter are either chosen over a grid of equally spaced values ranging from the maximum and the minimum of $\{W(\theta)\}$ observed, or on the basis of the quantiles or the variance of the observed values of $\{W(\theta)\}$. Adler et al. [2017] discuss this approach in the context of brain imaging, where a realization of the random field of interest (the brain) is available for each of the subjects under study (the patients). However, when different observations of the random field are not available, a Monte Carlo sample is needed. In this case, Result 6.2.1 provides a simplified solution to estimate $E[\phi(\mathcal{A}_{c_k})]$ and compute the Monte Carlo error associated with it. A discussion on the choice of the constants c_k in this setting is postponed to Section 6.4.2.

Notice that when $D = 1$, $\{W(\theta)\}$ is a random process and the search area is the interval $\Theta \equiv [\mathfrak{L}, \mathfrak{U}] \subset \mathbb{R}$. Hence $\mathcal{L}_0(\Theta) = 1$, and (6.7) specifies

$$E[\phi(\mathcal{A}_c)] = P(W(\mathfrak{L}) > c) + \mathcal{L}_1(\Theta)\rho_1(c) \quad (6.12)$$

where the second term in the right hand side of (6.12) is the so-called *expected Differential Topology (DT) characteristic of \mathcal{A}_c* [Worsley, 1994], and corresponds to the number of upcrossings of c by $\{W(\theta)\}$ [Chiu et al., 2013], i.e., $E[N_c]$ (see Chapter 3). Thus we write

$$E[N_c] = \mathcal{L}_1(\Theta)\rho_1(c), \quad (6.13)$$

hence, (3.4) in Result 3.1.2 is satisfied with $a(c) = \rho_1(c)$, and $b(\Theta) = \mathcal{L}_1(\Theta)$. Equation (6.14) follows from Result 3.1.3 and (6.13)

$$P\left(\sup_{\theta \in \Theta} \{W(\theta)\} > c\right) \approx P(W(\mathfrak{L}) > c) + \frac{\rho_1(c)}{\rho_1(c_0)} E[N_{c_0}] \quad \forall c_0 \geq c, c_0 \in \mathbb{R} \quad (6.14)$$

as $c \rightarrow \infty$. In general, when $D > 1$, the Lipschitz-Killing curvatures may be negative and thus, for small c , $E[\phi(\mathcal{A}_c)]$ may lead to an uninformative lower bound for $P(\sup_{\theta \in \Theta} \{W(\theta)\} > c)$. Conversely, when $D = 1$, the relationship in (6.13) guarantees that $\mathcal{L}_1(\Theta)$ is always non-negative and, on the basis of Result 6.2.1, so is its estimate $\mathcal{L}_1^*(\Theta)$. Thus, coherently with Result 3.1.3, when c is small the right

hand side of (6.14) leads to an upper bound for the left hand side.

6.2.1 LRT AND GLOBAL P-VALUES WHEN TESTING ON THE BOUNDARY

We consider a generalization of the mixture model in (1.14) where the nuisance parameter, $\boldsymbol{\theta}$, in $g(y, \boldsymbol{\theta})$ is allowed to be multidimensional. We are interested in testing (1.15) and (2.3); both tests are performed on the boundaries of the parameter space of η .

Denote with $\{K(\boldsymbol{\theta})\}$ the multivariate analogous of $\{K(\theta)\}$, i.e., we let $\{K(\boldsymbol{\theta})\}$ be a random field indexed by $\boldsymbol{\theta}$ and such that $K(\boldsymbol{\theta}) \sim \bar{\chi}_{01}^2$ for all $\boldsymbol{\theta} \in \Theta$. From Taylor and Worsley [2008] it follows that the EC densities, $\rho_d(c)$, of $\{K(\boldsymbol{\theta})\}$ are given by the sum of the EC densities of a χ_0^2 -random field and those of a χ_1^2 -random field, each multiplied by the respective mixture weight, i.e., 0.5. When $\Theta \subset \mathbb{R}^2$ as in Examples 1 and 2, (6.6) specifies as

$$E[\phi(\mathcal{A}_c)] = \frac{c^{\frac{1}{2}} e^{-\frac{c}{2}}}{(2\pi)^{\frac{3}{2}}} \mathcal{L}_2(\Theta) + \frac{e^{-\frac{c}{2}}}{2\pi} \mathcal{L}_1(\Theta) + \frac{P(\chi_1^2 > c)}{2} \mathcal{L}_0(\Theta) \quad (6.15)$$

where the functions of c multiplying $\mathcal{L}_0(\Theta), \dots, \mathcal{L}_2(\Theta)$ are the EC densities of a two-dimensional χ_1^2 random field (see Table (6.1)) divided by 2. Because the EC densities of a two-dimensional χ_0^2 random field evaluated at $c > 0$ are all zero, they do not contribute to (6.15).

In Section 6.4.2, we investigate via a suite of simulation studies the precision of (6.15) to approximate (6.2).

6.3 COMPUTING THE MEAN EULER CHARACTERISTIC VIA GRAPHS

We implement the approximation of the global p-value in Result 6.2.2 by estimating $E[\phi(\mathcal{A}_{c_k})]$, for c_1, \dots, c_D via a Monte Carlo simulation; this requires the evaluation of $\phi(\mathcal{A}_{c_k})$ for a sequence of realizations of $\{W(\boldsymbol{\theta})\}$. In this section we propose a convenient algorithm to achieve this goal.

To simplify notation, we assume that Θ is the cross product of the parameter spaces of the components of $\boldsymbol{\theta}$. Specifically, $\Theta = \Theta_1 \times \dots \times \Theta_D$, where Θ_d is the parameter space of component d of $\boldsymbol{\theta}$; the same reasoning easily applies when $\Theta \subset \Theta_1 \times \dots \times \Theta_D$ (e.g., Example 6.1 described in Section 6.4.1). In practice, we can only evaluate

Algorithm 1 Computing $\phi(\mathcal{A}_{c_k})$ via graphs

Input 1: Constant c_k .

Step 1: For all pairs $(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s)$ in $\tilde{\mathcal{A}}_{c_k}$ calculate the distance $d_\varphi(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s)$ in (6.16);

Step 2: construct the undirected graph $\mathcal{G}_k^D = (\tilde{\mathcal{A}}_{c_k}, E_k^D)$ where the edges E_k^D are allocated according to (6.17), with $d = D$;

Step 3: set $j = 1$;

Step 4: while $j < D$:

(i) set $d = D - j$;

(ii) obtain \mathcal{G}_k^d from \mathcal{G}_k^{d+1} by removing edges in E_k^{d+1} for which (6.17) does not hold;

(iii) count $|C_k^d|$ in \mathcal{G}_k^d via Eppstein et al. [2010];

(iv) $j=j+1$;

Step 5: calculate $\phi(\mathcal{A}_{c_k})$ via (6.18).

Output: Value of $\phi(\mathcal{A}_{c_k})$.

$\{W(\boldsymbol{\theta})\}$ on a finite set of values for $\boldsymbol{\theta}$. We do so by placing a grid of R_d points on Θ_d , for $d = 1, \dots, D$ and evaluating $\{W(\boldsymbol{\theta})\}$ at $\boldsymbol{\theta}_r = (\theta_{r1}, \dots, \theta_{rD})$ for $r = 1, \dots, R$, with $R = R_1 \times \dots \times R_D$, so that the evaluation points are the cross products of the component-wise grids. Finally, we let $\tilde{\Theta}_d$ be the ordered set of evaluation points of component d of $\boldsymbol{\theta}$ and Θ_\times be the full set of evaluation points of $\boldsymbol{\theta}$ over the cross product of $\tilde{\Theta}_1, \dots, \tilde{\Theta}_D$, i.e., $\Theta_\times = \{\boldsymbol{\theta}_r, r = 1, \dots, R\}$. For each constant c_k in Result 6.2.1, we define the excursion sets of $\{W(\boldsymbol{\theta}_r)\}$ above c_k to be the set of evaluation points $\tilde{\mathcal{A}}_{c_k} = \{\boldsymbol{\theta}_r \in \Theta_\times : W(\boldsymbol{\theta}_r) \geq c_k\}$, hence $\tilde{\mathcal{A}}_{c_k} \subseteq \Theta_\times$ provides a discretization of \mathcal{A}_{c_k} .

In order to compute $\phi(\mathcal{A}_{c_k})$ numerically, we consider a quadrilateral mesh^c of \mathcal{A}_{c_k} [Taylor and Worsley, 2008], i.e., the set of vertices composed of the points in $\tilde{\mathcal{A}}_{c_k}$ and the edges that connect them to form a partition of \mathcal{A}_{c_k} into D -dimensional hyperrectangles, and denoted by \mathcal{M}_k . Specifically, we consider the set of edges, E_k^1 , such that two vertices $\boldsymbol{\theta}_r$ and $\boldsymbol{\theta}_s$ in $\tilde{\mathcal{A}}_{c_k}$ are joined by an edge if and only if

^cFor simplicity, we limit our attention to the case of a quadrilateral mesh. However, our approach can be easily extended to any mesh involving regular polygons.

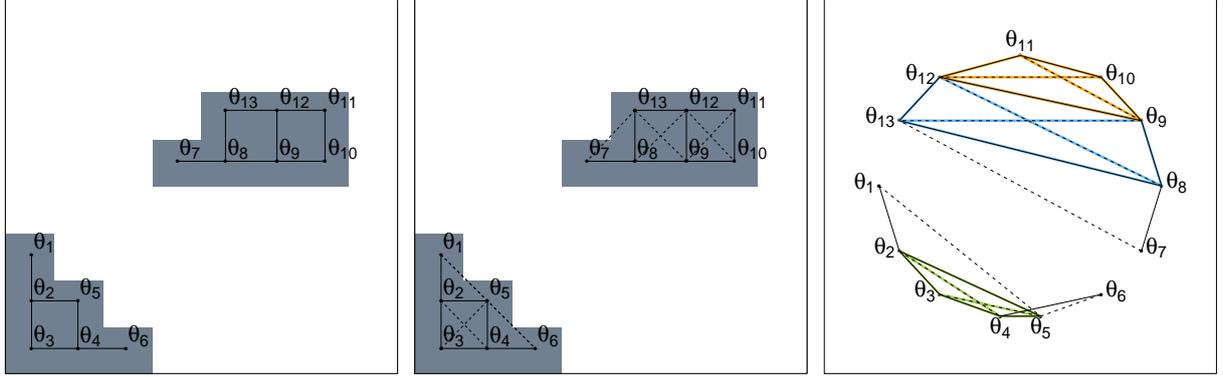


Figure 6.2: Left panel: quadrilateral mesh \mathcal{M}'_k of the excursion set \mathcal{A}_{c_k} (gray area), with set of vertices $\tilde{\mathcal{A}}_{c_k}$ (black dots) and edges E_k^1 allocated according to (6.16) (black solid segments) of unit length. Central panel: quadrilateral mesh \mathcal{M}'_k and diagonals of length $\sqrt{2}$ (black dashed segments). Right panel: graph $\mathcal{G}_k^2 = (\tilde{\mathcal{A}}_{c_k}, E_k^2)$ in which the three 4-dimensional cliques in C_k^2 are highlighted in orange, blue and green. As expected, each clique in \mathcal{G}_k^2 corresponds to a square in \mathcal{M}'_k .

$$d_\varphi(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s) = \sqrt{\sum_{d=1}^D (\varphi_d(r) - \varphi_d(s))^2} = 1, \quad (6.16)$$

where, $\varphi_d(r)$ is the index of component d of $\boldsymbol{\theta}_r$ within its (ordered) grid of evaluation points $\tilde{\Theta}_d$ and $d_\varphi(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s)$ is the Euclidean distance between the D indexes of the D components of $\boldsymbol{\theta}_r$ and $\boldsymbol{\theta}_s$ within the component-wise grids $\tilde{\Theta}_1, \dots, \tilde{\Theta}_D$. In \mathcal{M}_k , the lengths of the edges in E_k^1 are the Euclidean distances between $\boldsymbol{\theta}_r$ and $\boldsymbol{\theta}_s$, i.e., $d(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s) = \sqrt{\sum_{d=1}^D (\theta_{rd} - \theta_{sd})^2}$. In quadrilateral meshes involving only unit hypercubes $d(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s) = d_\varphi(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s)$.

We assume that Θ_\times is sufficiently dense to guarantee that \mathcal{A}_{c_k} is well approximated by \mathcal{M}_k . The EC is then calculated by alternatively adding and subtracting the number of d -dimensional hyperrectangles for $d = 0, \dots, D$ in \mathcal{M}_k [e.g., Adler, 2000]. In two dimensions for instance, the EC is obtained by counting the number of vertices, subtracting the number of edges and adding the number of rectangles [Worsley, 1995, Taylor and Worsley, 2008], e.g., Figure 6.1.

In order to ease computations in higher dimensions, one possible way to count the number of hyperrectangles of arbitrary large dimension d is summarized in Algorithm 1 and described below. The goal of Algorithm 1 is to construct graphs where the number of d -dimensional complete subgraphs (or cliques, to be defined soon) is equal to the number of d -dimensional hyperrectangles in \mathcal{M}_k . This can be done as follows.

For each constant c_k needed by Result 6.2.1, and for each dimension $d = 1, \dots, D$, consider an undirected unweighted graph, $\mathcal{G}_k^d = (\tilde{\mathcal{A}}_{c_k}, E_k^d)$, with vertices $\tilde{\mathcal{A}}_{c_k}$ and edges E_k^d such that two vertices θ_r and θ_s are joined by an edge if and only if

$$1 \leq \mathbf{d}_\varphi(\theta_r, \theta_s) \leq \sqrt{d} \quad (6.17)$$

where \sqrt{d} corresponds to the length of the longest diagonal of a d -dimensional unit hypercube.

A general graph $\mathcal{G} = (V, E)$ has a clique of dimension Q if there exists a subset of Q vertices in V such that every pair of distinct vertices of the subset are connected by an edge. We denote the set of all 2^d -dimensional cliques in \mathcal{G}_k^d by C_k^d . The distance between points in $\tilde{\mathcal{A}}_{c_k}$ does not affect the enumeration of the hyperrectangles in \mathcal{M}_k . Specifically, since the allocation of the edges E_k^1 only depends on the indexes $\varphi_d(r)$ of the θ_{rd} within $\tilde{\Theta}_d$, for $d = 1, \dots, D$, the number of d -dimensional hyperrectangles in \mathcal{M}_k is equal to the number of d -dimensional unit hypercubes in a “unit” mesh, denoted by \mathcal{M}'_k , with vertices $\tilde{\mathcal{A}}_{c_k}$ and edges E_k^1 of unit length. It follows that the 2^d vertices of each clique in C_k^d is a subset of points in $\tilde{\mathcal{A}}_{c_k}$ which are at least one unit, and at most \sqrt{d} , apart one another. By construction, this implies that each clique in C_k^d corresponds to a unit d -dimensional hypercube in \mathcal{M}'_k , which in turn corresponds to a d -dimensional hyperrectangle in \mathcal{M}_k . For illustrative purposes, in Figure 6.2 we give an example in two dimensions, where for simplicity the points θ_r are equally spaced over unit intervals in each $\tilde{\Theta}_d$, $d = 1, 2$, and thus $\mathcal{M}_k = \mathcal{M}'_k$. Notice that the main difference between the mesh \mathcal{M}_k (or \mathcal{M}'_k) and the graph \mathcal{G}_k^D is that the former depends on the position of the points in $\tilde{\mathcal{A}}_{c_k}$ over Θ and their distance; whereas the latter only accounts for their connectivity.

Therefore, in general terms, we can compute $\phi(\mathcal{A}_{c_k})$ as

$$\phi(\mathcal{A}_{c_k}) = \sum_{d=0}^D (-1)^d |C_k^d| \quad (6.18)$$

$$= |\tilde{\mathcal{A}}_{c_k}| - |E_k| + \sum_{d=2}^D (-1)^d |C_k^d| \quad (6.19)$$

where $|\cdot|$ is the cardinality of the set considered. Equation (6.19) follows from (6.18) since by construction $\mathcal{G}_k^0 = \tilde{\mathcal{A}}_{c_k}$, \mathcal{G}_k^1 is the unweighted graph with the same vertices and edges of \mathcal{M}_k and \mathcal{M}'_k ; thus $|C_k^0| = |\tilde{\mathcal{A}}_{c_k}| = \sum_{r=1}^R \mathbb{1}_{\{w(\theta_r) > c_k\}}$ and $|C_k^1| = |E_k^1| = \sum_{r=1}^R \sum_{s=1}^R \mathbb{1}_{\{\mathbf{d}_\varphi(\theta_r, \theta_s) = 1\}}$.

Naively, computing $|C_k^d|$ by sequentially considering each subset of $\tilde{\mathcal{A}}_{c_k}$ of size 2^d requires a complexity $O(|\tilde{\mathcal{A}}_{c_k}|^{2^D} 4^D)$ to evaluate (6.18), a massive computation load unless D is quite small. The advantage of converting the hyperrectangles enumeration problem into a clique-finding problem is that several efficient algorithms exist to address this challenge in near-optimal time [e.g., Bron and Kerbosch, 1973, Johnston, 1976, Eppstein et al., 2010]. In our implementations in Section 2.2, we use the algorithm proposed by Eppstein et al. [2010], and implemented in the R function `cliques` in the `igraph` package [Csardi and Nepusz, 2006]. Specifically, Eppstein et al. [2010] propose a variation of the Bron-Kebosch algorithm for sparse graphs^d where the running time is of $O(h|\tilde{\mathcal{A}}_{c_k}|^{\frac{h}{3}})$, with $h = 2^D - 1$. This is particularly convenient in our context where the constants c_k can be chosen arbitrarily to reduce both the size of the graph and its sparsity. Hence in Algorithm 1 we recommend a top-down approach where \mathcal{G}_k^D is constructed first, and the constants c_k can be adequately adjusted between Step 2 and Step 3 in order to increase sparsity in \mathcal{G}_k^D . The graphs \mathcal{G}_k^d , for $d = 0, \dots, D - 1$, are obtained subsequently by removing edges for which (6.17) is not satisfied as d decreases. An additional advantage of this approach is that \mathcal{G}_k^D provides a simple two-dimensional representation of the D -dimensional excursion sets \mathcal{A}_{c_k} .

Finally, for adequate choices of c_k (see Section 6.4.2), Monte Carlo estimates of $E[\phi(\mathcal{A}_{c_k})]$, namely $E[\widehat{\phi(\mathcal{A}_{c_k})}]$, can be obtained by computing Algorithm 1 over a small set of Monte Carlo replicates of $\{W(\theta_r)\}$ and averaging over the values $\phi(\mathcal{A}_{c_k})$ calculated on each replicate. (The reader is referred to Section 6.4.2 for a discussion on the accuracy of $E[\widehat{\phi(\mathcal{A}_{c_k})}]$.) Consequently, we can approximate the right hand sides of (6.11) with

$$\mathcal{L}_0(\Theta)P(W(\theta_1) > c) + \sum_{j=1}^D \widehat{\mathcal{L}_d^*(\Theta)} \rho_d(c) \quad (6.20)$$

where $\widehat{\mathcal{L}_d^*(\Theta)}$ are the solution of the system of equation in (6.10) with $E[\phi(\mathcal{A}_{c_k})]$ in the left hand sides of each equation replaced by their Monte Carlo estimates $E[\widehat{\phi(\mathcal{A}_{c_k})}]$.

^dThe Bron-Kebosch algorithm finds all maximal cliques in a graph, i.e., it lists all subsets of vertices for which each pair is connected by an edge, and such that, if an additional vertex is added to any of these subsets the complete connectivity is violated [Bron and Kerbosch, 1973].

6.4 NUMERICAL RESULTS

6.4.1 MOTIVATING EXAMPLES

In this section we show how TOHM can be used in the context of feature detection in images, comparison of non-nested models characterized by multidimensional parameters and a logistic regression with a break point and change of trend. We refer to these examples with Example 6.1, 6.2 and 6.3, respectively; notice that in all of them the multidimensional parameter which characterizes the structural change in data distribution, and is not identifiable under H_0 , is denoted with θ .

Example 6.1. We consider a dark matter search in the sky where the goals are (i) to assess the presence of a photons emission due to a dark matter source in addition to background photons, and (ii) to identify the location at which maximum evidence in favor of the suspected source is achieved. Specifically, we consider realistic simulations of the Fermi Large Area Telescope (LAT) where the astrophysical background is uniformly distributed over a disc in the sky of 30° radius and centered at (195RA, 28DEC), which corresponds to our search region Θ . In our simulations the dark matter source is modeled as a bivariate Gaussian located at $(\theta_1, \theta_2) = (174.952, 37.986)$. Realistic representations of the systematic errors, as well as the calibration of the detector, were included in the simulation obtained via the *gtobssim* package^e. This set up led to 51,098 background events and 39 dark matter events. We specify the model of interest as

$$(1 - \eta) \frac{1}{\lambda(\Theta)} + \eta \frac{1}{k_{\theta_1\theta_2}} \exp \left\{ -\frac{1}{0.02} \left[\left(\frac{x - \theta_1}{\theta_1} \right)^2 + \left(\frac{y - \theta_2}{\theta_2} \right)^2 \right] \right\} \quad (6.21)$$

where $x \in [165; 195]$, $y \in [28 - \sqrt{30^2 - (x - 195)^2}; 28 + \sqrt{30^2 - (x - 195)^2}]$, $\eta \in [0, 1]$ is the intensity of the dark matter emission, $\theta = (\theta_1, \theta_2)$ is the location of the emission over the disc Θ , $k_{\theta_1\theta_2}$ is a normalizing constant and $\lambda(\Theta)$ is the area of Θ . We assess the presence of the signal by testing (1.15).

Example 6.2. We consider the *Compressive strength and strain of maize seeds dataset* available in the R package `goft` [Gonzalez-Estrada and Villasenor-Alva, 2016]. The dataset records the compression strength in Newtons of 90 seeds and the goal is to choose between a gamma and a log-normal distribution for the data.

^e<http://fermi.gsfc.nasa.gov/ssc/data/analysis/software>

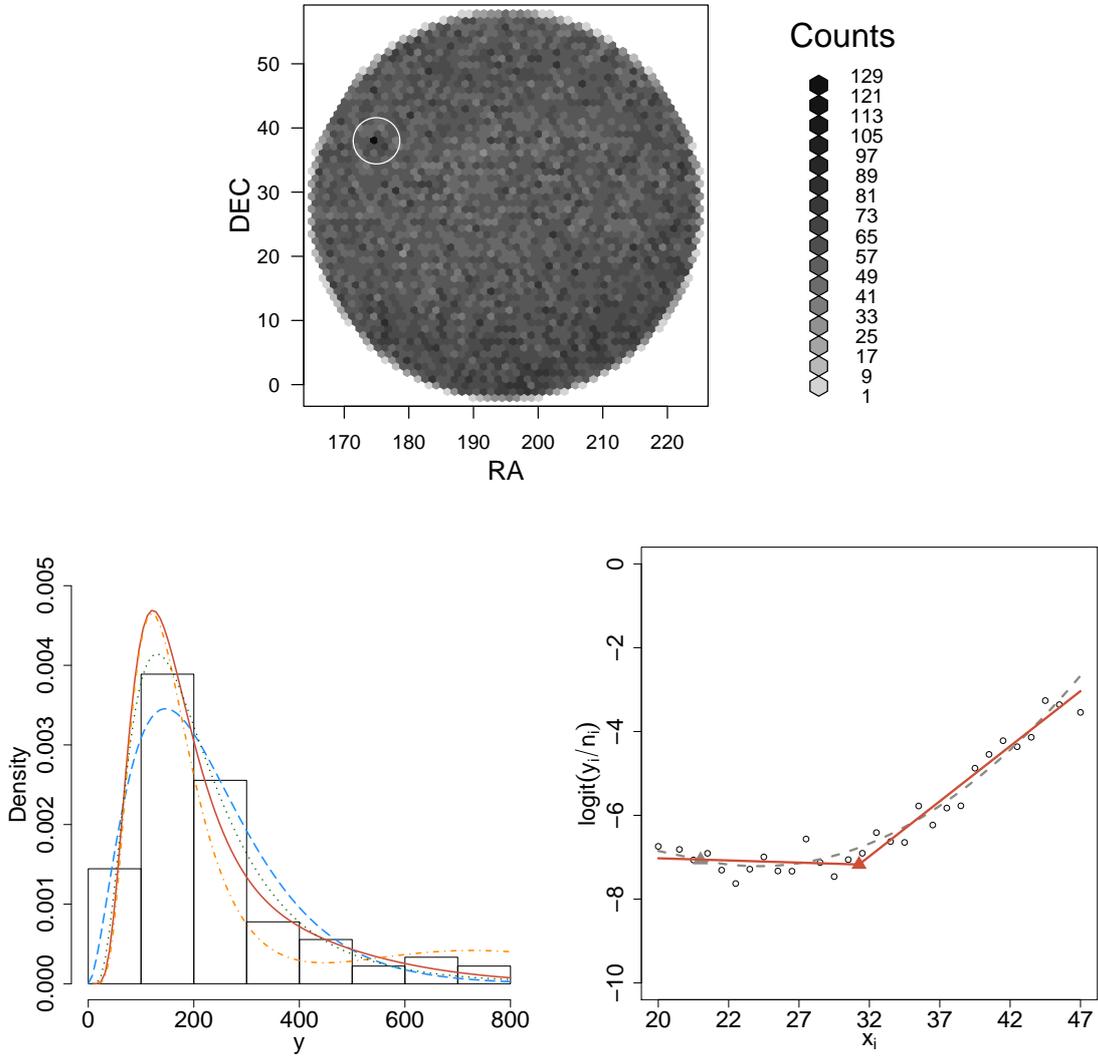


Figure 6.3: Data, null and fitted models. Top panel: 2D histogram of the Fermi-LAT realistic data simulation for Example 6.1. The white circle indicates the location at which the LRT-process achieves its maximum, i.e., $\hat{\theta} = (175, 38)$ with estimated intensity $\hat{\eta}_{\hat{\theta}} = 0.001$. Bottom left panel: histogram of the maize seeds strength data for Example 6.2. The null model in (1.15) (blue dashed curve) is fitted as a gamma distribution with $(\hat{\tau}, \hat{\gamma}) = (2.762, 83.007)$. The null model in (2.3) (green dotted line) is fitted as a log-normal distribution with $(\hat{\mu}, \hat{\sigma}) = (5.243, 0.614)$. The alternative model when testing (1.15) (red solid line) is fitted with $(\hat{\eta}_{\hat{\mu}\hat{\sigma}}, \hat{\gamma}_{\hat{\mu}\hat{\sigma}}, \hat{\tau}_{\hat{\mu}\hat{\sigma}}, \hat{\mu}, \hat{\sigma}) = (0.783, 4.820, 88.742, 5.041, 0.5)$. Finally the alternative model when testing (2.3) (chained orange line) is fitted with $(\hat{\eta}_{\hat{\gamma}\hat{\tau}}, \hat{\gamma}, \hat{\tau}, \hat{\mu}_{\hat{\gamma}\hat{\tau}}, \hat{\sigma}_{\hat{\gamma}\hat{\tau}}) = (0.741, 9.816, 83.061, 5.014, 0.472)$. Bottom right panel: Down syndrome data, the model in (6.23) selected by THOM is a break point logistic regression with linear trend (red solid lines) i.e., $\tilde{\theta} = (\tilde{\theta}, \tilde{\alpha}) = (31.265, 1)$, with a break point at $\tilde{\theta} = 31.265$ (red triangle). For comparison, a break point logistic regression with change of trend from linear to quadratic (gray dashed line) is also fitted while fixing $\alpha = 2$. In this case the breakpoint occurs at $\tilde{\theta} = 20$ (gray triangle).

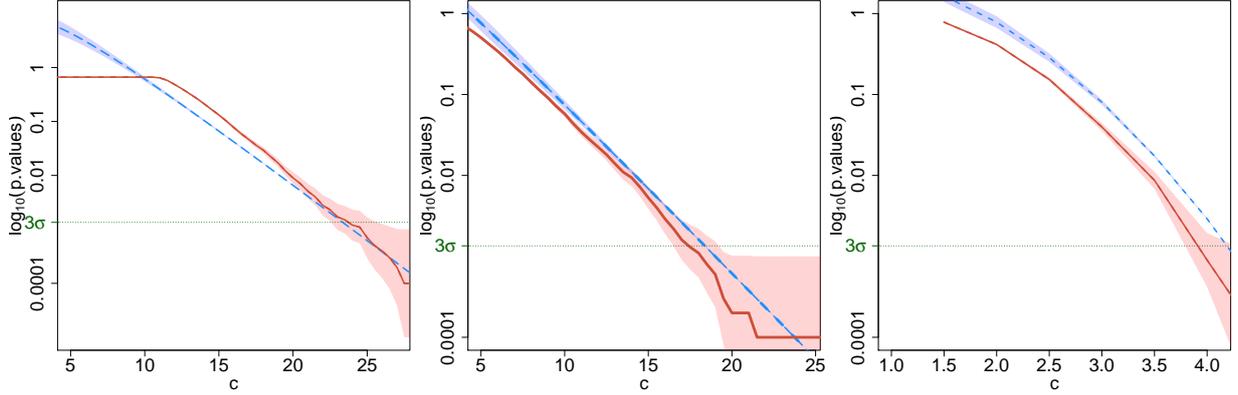


Figure 6.4: Estimated approximations in (6.20) (blue dashed line), Monte Carlo estimates of $P(\sup_{\theta \in \Theta} \{W(\theta)\} > c)$ (red solid line) in \log_{10} -scale, and Monte Carlo Errors (pink areas) for increasing values of the threshold c , for Example 6.1 (left panel), Example 6.2 (central panel) and Example 6.3 (right panel). Monte Carlo errors associated with $E[\widehat{\phi}(\mathcal{A}_{c_k})]$ in (6.20) are plotted as gray areas. In each plot the sample size of each Monte Carlo dataset is 100,000, the number of Monte Carlo replicates used to obtain the Monte Carlo p-values is 10,000, whereas the quantities $E[\widehat{\phi}(\mathcal{A}_{c_k})]$ have been estimated over a separate set of 100 Monte Carlo replicates.

Hence the we consider the comprehensive model

$$(1 - \eta) \frac{e^{-y/\tau} y^{\gamma-1}}{k_{\tau\gamma}} + \eta \frac{\exp\left\{-\frac{\ln y - \mu}{2\sigma^2}\right\}}{y k_{\mu\sigma}}, \quad (6.22)$$

where $\eta \in [0, 1]$, $\gamma > 0$, $\tau > 0$, $k_{\tau\gamma}$ and $k_{\mu\sigma}$ are normalizing constants. In order to ease our computation, we let $y \in (0, 1000]$. In this case the parameter which is present only under the alternative is $\theta = (\mu, \sigma)$ when testing (1.15) and $\theta = (\gamma, \tau)$ when testing (2.3).

Example 6.3. We refer to the *Down Syndrome dataset* introduced in Example 3.2. However, in contrast to Example 3.2, here we allow for a change of trend after the break point. Specifically, we allow for a quadratic trend, a change in the slope of the linear trend or a break due to a change of the intercept, i.e., the model of interest is

$$E\left[\log\left(\frac{\pi_i}{1 - \pi_i}\right)\right] = \phi_1 + \phi_2 x_i + \xi(x_i - \theta)^\alpha \mathbb{1}_{\{x_i \geq \theta\}} \quad \text{for } i = 1, \dots, n \quad (6.23)$$

where $\pi_i = P(Y_i = 1)$, $x \in \mathbb{R}$, $\mathbb{1}_{\{\cdot\}}$ is the indicator function, $\theta = (\theta, \alpha)$, with $\alpha \in \{0, 1, 2\}$, we let $\Theta \equiv [-12, 12] \times \{0, 1, 2\}$, and we test (3.22).

For Examples 6.1 and 6.2 the tests in (1.15) and/or (2.3) are performed via the classical LRT. Hence, we consider the random field $\{K(\theta)\}$ with components $K(\theta) \sim$

$\bar{\chi}_{01}^2$. In Example 6.3, we consider the signed-root-LRT, $Q_n(\boldsymbol{\theta})$, in (3.23). For each $\boldsymbol{\theta}$ and under H_0 , $Q_n(\boldsymbol{\theta})$ is asymptotically normally distributed with mean-zero and unit variance. Further, by exploiting the asymptotic equivalence of $Q_n(\boldsymbol{\theta})$ and the normalized Score function [Davies, 1977, Moran, 1970], it can be shown [Pilla et al., 2005] that the supremum of the random field $\{Q_n(\boldsymbol{\theta})\}$ converges in distribution to the supremum of a mean zero and unit variance Gaussian random field.

The data for Examples 6.1-6.3 are plotted in Figure 6.3.

6.4.2 GOODNESS OF THE APPROXIMATIONS AND CHOICE OF c_k

Our first task is to assess the validity of the approximation of (6.2) via (6.20), as $c \rightarrow \infty$.

In the plots in Figure 6.4 we show as red dashed lines the Monte Carlo estimates of $P(\sup_{\boldsymbol{\theta} \in \Theta} \{W(\boldsymbol{\theta})\} > c)$ obtained using 10,000 data sets simulated under the null model. In order to guarantee that the asymptotic distribution of the test statistics considered is achieved, we simulate at each replicate 100,000 events; their Monte Carlo errors are given by the pink areas. These are compared with the approximation in (6.20) plotted as blue dashed lines as c increases (x -axis). We use a set of 100 Monte Carlo replicates, again each of size 100,000, to estimate the quantities $E[\phi(\mathcal{A}_{c_k})]$ used in (6.10) to obtain each $\mathcal{L}_d^*(\Theta)$ in (6.9) with $\phi(\mathcal{A}_{c_k})$ computed via Algorithm 1 at each replicate.

For Example 6.1 (left panel of Figure 6.4), we considered a grid of size $R = 2821$ over the 30 degree radius circular search region centered at (195RA, 28DEC). Since in this case Θ is given by a disc, its EC is one and thus $\mathcal{L}_0(\Theta) = 1$. In order to estimate $\mathcal{L}_1(\Theta)$ and $\mathcal{L}_2(\Theta)$ we consider $c_1 = 1$ and $c_2 = 8$, which lead to an accurate approximation of $P(\sup_{\boldsymbol{\theta} \in \Theta} \{W(\boldsymbol{\theta})\} > c)$, with $\widehat{\mathcal{L}_1^*(\Theta)} = -244.053$ and $\widehat{\mathcal{L}_2^*(\Theta)} = 644.244$.

For Example 6.2 in the central panel of Figure 3.4, we define a grid of size $R = 2500$ over the square $[1, 10] \times [0.2, 5]$. Again $\mathcal{L}_0(\Theta) = 1$, and we chose $c_1 = 2$ and $c_2 = 3$. The resulting estimates for the Lipschitz-Killing curvatures are $\widehat{\mathcal{L}_1^*(\Theta)} = 30.11037$ and $\widehat{\mathcal{L}_2^*(\Theta)} = 30.52665$, which in this case also lead to a good approximation of $P(\sup_{\boldsymbol{\theta} \in \Theta} \{W(\boldsymbol{\theta})\} > c)$.

Finally, the right panel of Figure 6.4 shows the goodness of the approximation provided by (6.20) for Example 6.3. The parameter space Θ corresponds to $[-12, 12] \times [0, 2]$, and we let $R=150$ as in practice we only allow values of α equal to 0, 1 and

Example	$\tilde{\theta}$	P-value (significance)	Error (significance interval)
Example 6.1	$(\tilde{\theta}_1, \tilde{\theta}_2)$ (175, 38)	$1.092 \cdot 10^{-26}$ (10.629 σ)	$9.272 \cdot 10^{-28}$ [10.621 σ ; 10.637 σ]
Example 6.2 $H_0 : \eta = 0$ vs $H_0 : \eta > 0$	$(\tilde{\mu}, \tilde{\sigma})$ (5.041, 0.5)	0.036 (1.801 σ)	0.012 [1.663 σ ; 1.988 σ]
Example 6.2 $H_0 : \eta = 1$ vs $H_0 : \eta < 1$	$(\tilde{\gamma}, \tilde{\tau})$ (4.898, 82.653)	0.624 (0.00 σ)	0.096
Example 6.3	$(\tilde{\theta}, \tilde{\alpha})$ (31.265, 1)	$8.611 \cdot 10^{-30}$ (11.276 σ)	$2.494 \cdot 10^{-30}$ [11.254 σ ; 11.306 σ]

Table 6.2: TOHM p-values computed via (6.20), σ -significance and respective errors.

2. Selecting $c_1 = 0.5$ and $c_2 = 1$, we obtain $\widehat{\mathcal{L}}_1^*(\Theta) = 16.724$ and $\widehat{\mathcal{L}}_2^*(\Theta) = 23.291$ which lead to a satisfactory approximation for the global p-value.

Since Result 6.2.1 and Result 6.2.2 hold for any choice of c_k , $k = 1, \dots, D$, it is convenient to choose them sufficiently small so that the excursion sets \mathcal{A}_{c_k} are composed by a reasonably high number of connected components; this reduces the size of the Monte Carlo simulation required to accurately estimate the quantities $E[\phi(\mathcal{A}_{c_k})]$. Hence, the thresholds c_k should be chosen small enough that the excursion sets \mathcal{A}_{c_k} are non-empty with high probability. Additionally, since both the size and the sparsity of the graphs \mathcal{G}_k^D affects the running time of Algorithm 1, c_k should be selected accordingly. These aspects can be assessed with a sensitivity analysis. Specifically, for a given c_k , \mathcal{G}_k^D allows a two-dimensional visualization of the D -dimensional mesh \mathcal{M}_k , and thus after Step 2 in Algorithm 1, c_k can be increased to increase sparsity and decrease the size of \mathcal{G}_k^D before proceeding with Steps 3-5.

In principle, the choice of c_1, \dots, c_D should also take into account the possibility that the ECs, $\phi(\mathcal{A}_{c_k})$, for different values of k could be correlated with one another. However, since we are interested in $c \rightarrow \infty$, the Monte Carlo error associated with (6.20) become extremely small as c increases, and this is true even when, as in Figure 6.4, the quantities $\phi(\mathcal{A}_{c_k})$ have been computed on the same set of Monte Carlo simulations for each c_k considered.

6.4.3 DATA ANALYSIS

We calculated the TOHM p-value in (6.20) and respective significance for Examples 6.1, 6.2 and 6.3. The results are summarized in Table 6.2.

In Example 6.1, we performed $R = 2821$ tests over our circular search region centered at (195RA, 28DEC). In our realistic simulation, the true dark matter emission was located at (174.952RA, 37.986DEC) and the LRT-process used in TOHM achieves its maximum at $\tilde{\theta} = (175\text{RA}, 38\text{DEC})$ with about 10σ significance. Notice that our original dataset includes 51,098 background events and only 39 dark matter events; hence the procedure appears to be particularly powerful even in presence of a low signal-to-noise ratio. The identified location is plotted as a white circle in the upper panel of Figure 6.3.

In Example 6.2, we set $R = 2500$ when testing (1.15) and the gamma model is rejected at a 0.05 significance level by the THOM p-value. Whereas, when testing (2.3), the log-normal model cannot be rejected; the resulting p-value is 0.624. Thus, the log-normal model is selected for the maize seeds strength data, and the maximum is achieved at $\tilde{\mu} = 5.004$ and $\tilde{\sigma} = 0.633$. The log-normal fitted model is plotted in the bottom left panel of Figure 6.3 as a red solid line.

Finally in Example 6.3, when testing (3.22) $R = 150$ times, the global p-value computed via (6.20) provides strong evidence ($\sim 11\sigma$) in favor of a linear trend ($\tilde{\alpha} = 1$) with a break point at $\tilde{\theta} = 31.265$. Hence we expect the risk of giving birth to a child with down syndrome to increase when the mother is 31 years old or older. The model selected is displayed as a red solid line in the bottom right panel of Figure 6.3, with the break-point indicated by a red triangle. For the sake of comparison, we also plot the fitted model when allowing a quadratic trend ($\alpha = 2$) with a break point selected by the procedure at $\tilde{\theta} = 20$.

CONCLUSION

The main findings collected in this thesis can be summarized as follow.

In Chapter 1 we review several approaches proposed in both the statistics and the physics literature to test hypotheses under standard and non-standard regularity conditions. These include tests conducted on the boundary of the parameter space [Chernoff, 1954], tests where a nuisance parameter is present only under the alternative [Davies, 1987, Gross and Vitells, 2010, Pilla and Loader, 2005, Pilla et al., 2005] and classical multiple hypothesis testing [Bonferroni, 1935, 1936]. This allows us to introduce the background necessary for the developments and the comparisons discussed in the subsequent chapters.

In Chapter 2 we show that we can select between non-nested models by means of a comprehensive model of which the two models under comparison are special cases. The problem is then reduced into a test of hypothesis for nested models where the parameter of interest, η , is tested on the boundary of its parameter space while a nuisance parameter, θ , is present only under the alternative. The test is performed by adequately adjusting the method of Gross and Vitells [2010] to correct for the look-elsewhere effect on the basis of the results in Chernoff [1954]. This approach is implemented in the context of indirect searches for dark matter which mainly motivate the developments in this chapter.

In Chapter 3 we formalize the methodological framework of TOHM in one-dimension and show how EVT and Monte Carlo methods can be combined to efficiently approximate the supremum of a stochastic process and obtain global p-values. This generalizes the solution of Gross and Vitells [2010] beyond the LRT and χ^2 case and clarifies the conditions necessary for these results to hold. We also provide guidelines and graphical tools to validate the adequacy of the grid resolution and select the threshold c_0 used to effectively reduce the Monte Carlo error.

In Chapter 4 we identify scenarios where TOHM leads to the same inference of MHT via Bonferroni, and we introduce a simple plot to assess when equivalence among the two hold. On the other hand, we show that when the two testing procedures lead to different results, the TOHM bound is typically less conservative when the number R of (sub-)tests conducted is large and, unlike Bonferroni, yields valid inference.

Conversely, when R is small, Bonferroni is less conservative than TOHM and it is therefore preferred in such cases.

In Chapter 5 we show that, for small samples sizes, the testing procedure of Pilla et al. [2005] and Pilla and Loader [2005] may produce a higher number of false detections than expected. Conversely, the method of Gross and Vitells [2010] adheres to the expected probability of type I error even for small sample sizes, thus it is preferred to obtain global p-values when n is small. We also discuss how, for large R , the computational effort can be reduced by selecting sequentially among different testing procedures while maintaining the desired statistical properties.

Finally, in Chapter 6 we generalize the framework of Chapter 3 to multiple dimensions and we obtain a computational solution which combines the probabilistic constructs of differential geometry for random fields [Taylor and Worsley, 2008] with simple results in graph theory to efficiently approximate the distribution of the supremum of a random field. This implicitly extends the methods proposed by Davies [1977, 1987], Gross and Vitells [2010] and Vitells and Gross [2011] to test hypothesis when a multidimensional parameter is present only under the alternative.

The results of Chapter 3 and Chapter 6 are of particular importance as they allow us to establish a unified framework to perform inference under non-standard regularity conditions. Specifically, we show how TOHM can be used to detect an unexpected mode in the data, compare non-nested model, detect a change point in a trend, and deal with several other situations.

Stringent significance requirements play a critical role in both the theory and the practical implementation of TOHM. Hence, our setup is particularly well suited for searches in high energy physics, where the significance level necessary to claim a discovery is of at least 5σ (see Table 1.1). However, in light of the recent “p-value crisis”, culminated with the *Journal Basic and Applied Social Psychology* banning the use of the p-value in future submissions [Wasserstein and Lazar, 2016, Leek and Peng, 2015], stringent significance criteria may become more popular in other scientific communities.

Despite its general character, TOHM suffers from several limitations which offer opportunities for future development. Specifically, from a computational perspective, TOHM requires a simulation stage in which the test statistic considered is evaluated over a fine grid on the parameter space, for each of the (few) Monte Carlo replicates under H_0 . Hence, the computation can become particularly heavy for large dimensional parameters. One could naively circumvent the problem by reducing

the resolution of the grid or limiting the test to a small region of the parameter space. We aspire to a more sophisticated solution which would incorporate an importance sampling-like approach to select regions of the parameter space where the grid evaluation is performed.

From a methodological perspective, TOHM targets situations where only one signal is expected to be identified over the search range (e.g., one bump, one break point in the regression line etc). Thus, a natural question is: *can we utilize the framework of TOHM to identify a few (but still ≥ 1) signals under stringent significance requirements?* The statistical tool to be developed in Algeri [2018] addresses this question via a multiple hypothesis testing approach where the number of tests to be conducted is reduced via extreme value theory. The major advantage of this approach is that it does not require the specification of the expected number of signals and thus, provides a comprehensive tool which simultaneously solves important long-standing problems such as detecting the number of components in a mixture or identifying multiple bumps in the data, and it is applicable in one or more dimensions. The fundamental problem of detecting dark matter sources in the Galactic Center [Daylan et al., 2017] offers the main motivation for this work. Preliminary simulation results are coherent with the outline of the theory which we have already developed. More time is needed for the proof of the intermediate theoretical steps, as well as for the application on real data, the development of a validation method, and the assessment of the power of the procedure.

It is important to point out that TOHM relies on the assumptions that the parametric models provided by theoreticians reliably represent both the true underlying process and the systematic uncertainties associated with the detector, instrumental resolution etc. In several physics and astronomy problems, however, these assumption fail in practice. Examples include the direct search for dark matter [Undagoitia and Rauch, 2015], the discrimination of the neutrino mass hierarchy [Capozzi et al., 2015] and searching for rare decays of b-hadrons [Koppenburg et al., 2016]. In the absence of valid theoretical models and/or strong prior knowledge of the systematics effects and/or sufficient computational resources to estimate them, a reliable data-driven solution has the potential for major impact in the physics community. To accomplish this, we will develop the core of the statistical solution in a non-parametric framework. Specifically, given a parametric model, which incorporates all the (limited) knowledge available with respect to the background process, the data are used to update the input model, for example, using the modern approach of Mukhopadhyay (2016). As an additional advantage, the methodology we intend

to propose does not require the specification of a model for the signal.

Finally, we aim to develop an R package to automatically implement all the methods discussed in this thesis. This software solution has the potential of being an indispensable element in the “tool-box” of any experimental physicist or astronomer, or really any applied scientist.

References

- G. Aad et al. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, 716(1):1–29, 2012.
- R.J. Adler. *The geometry of random fields*, volume 62. Siam, 1981.
- R.J. Adler. On excursion sets, tube formulas and maxima of random fields. *Annals of Applied Probability*, pages 1–74, 2000.
- R.J. Adler and J.E. Taylor. *Random fields and geometry*. Springer Science & Business Media, 2009.
- R.J. Adler et al. Estimating thresholding levels for random fields via euler characteristics. *arXiv preprint arXiv:1704.08562*, 2017.
- S. Algeri. R package nonnest. 2015. URL <http://wwwf.imperial.ac.uk/~sa2514/Research.html>.
- S. Algeri. Dimension reduction in multiple hypothesis testing: a solution to the k signals problem. *Working paper*, 2018.
- S. Algeri and D.A. van Dyk. Testing one hypothesis multiple times. *arXiv preprint arXiv:1701.06820*, 2017.
- S. Algeri and D.A. van Dyk. Testing one hypothesis multiple times: The multidimensional case. *In Preparation*, 2018.
- S. Algeri, J. Conrad, and D.A. van Dyk. A method for comparing non-nested models with application to astrophysical searches for new physics. *Monthly Notices of the Royal Astronomical Society: Letters*, 458(1):L84–L88, 2016a.
- S. Algeri et al. On methods for correcting for the look-elsewhere effect in searches for new physics. *Journal of Instrumentation*, 11(12):P12010, 2016b.
- D.W.K. Andrews and W. Ploberger. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica: Journal of the Econometric Society*, pages 1383–1414, 1994.

- A.C. Atkinson. A method for discriminating between models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 323–353, 1970.
- W. B. Atwood et al. The large area telescope on the fermi gamma-ray space telescope mission. *The Astrophysical Journal*, 697(2):1071, 2009.
- Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, pages 289–300, 1995.
- L. Bergström, P. Ullio, and J.H. Buckley. Observability of γ rays from dark matter neutralino annihilations in the milky way halo. *Astroparticle Physics*, 9(2):137–162, 1998.
- S.M. Berman. Limit theorems for the maximum term in stationary sequences. *The Annals of Mathematical Statistics*, pages 502–516, 1964.
- C.E. Bonferroni. Il calcolo delle assicurazioni su gruppi di teste. In *Studi in Onore del Professore Salvatore Ortu Carboni*, pages 13–60. Rome, 1935.
- C.E. Bonferroni. Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, 8:3–62, 1936.
- C. Bron and J. Kerbosch. Algorithm 457: finding all cliques of an undirected graph. *Communications of the ACM*, 16(9):575–577, 1973.
- F. Capozzi, E. Lisi, and A. Marrone. Neutrino mass hierarchy and precision physics with medium-baseline reactors: Impact of energy-scale and flux-shape uncertainties. *Physical Review D*, 92(9):093011, 2015.
- S. Chatrchyan et al. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Physics Letters B*, 716(1):30–61, 2012.
- H. Chernoff. On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics*, pages 573–578, 1954.
- S.N. Chiu et al. *Stochastic geometry and its applications*. John Wiley & Sons, 2013.
- J. Conrad. Statistical issues in astrophysical searches for particle dark matter. *Astroparticle Physics*, 62:165–177, 2015.

- R. Cousins et al. Spin discrimination of new heavy resonances at the lhc. *Journal of High Energy Physics*, 2005(11):046, 2005.
- D.R. Cox. Tests of separate families of hypotheses. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 105–123, 1961.
- D.R. Cox. Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 406–424, 1962.
- D.R. Cox. A return to an old paper:tests of separate families of hypotheses. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(2):207–215, 2013.
- H. Cramér and M.R. Leadbetter. *Stationary and related stochastic processes: Sample function properties and their applications*. Courier Corporation, 2013.
- G. Csardi and T. Nepusz. The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695(5):1–9, 2006.
- R.B. Davies. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 64(2):247–254, 1977.
- R.B. Davies. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 74(1):33–43, 1987.
- A.C. Davison. *Statistical models*, volume 11. Cambridge University Press, 2003.
- T. Daylan, S.K.N. Portillo, and D.P. Finkbeiner. Inference of unresolved point sources at high galactic latitudes using probabilistic catalogs. *The Astrophysical Journal*, 839(1):4, 2017.
- M. Della Negra, P. Jenni, and T.S. Virdee. Journey in the search for the higgs boson: The atlas and cms experiments at the large hadron collider. *Science*, 338(6114):1560–1568, 2012.
- B. Efron. *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*, volume 1. Cambridge University Press, 2012.
- B. Efron and R.J. Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.
- J. Ellis, M.K. Gaillard, and D.V. Nanopoulos. A historical profile of the higgs boson. *arXiv:1201.6045*, 2012.

- D. Eppstein et al. *Listing All Maximal Cliques in Sparse Graphs in Near-Optimal Time*, pages 403–414. Springer Berlin Heidelberg, 2010.
- M. Falk, J. Hüsler, and R.-D. Reiss. *Laws of small numbers: extremes and rare events*. Springer Science & Business Media, 2010.
- G.J. Feldman and R.D. Cousins. Unified approach to the classical statistical analysis of small signals. *Physical Review D*, 57(7):3873, 1998.
- E. Gonzalez-Estrada and J.A. Villasenor-Alva. *goft: Tests of Fit for some Probability Distributions*, 2016. URL <https://CRAN.R-project.org/package=goft>. R package version 1.3.1.
- E. Gross and O. Vitells. Trial factors for the look elsewhere effect in high energy physics. *The European Physical Journal C*, 70(1-2):525–530, 2010.
- B.E. Hansen. Inference when a nuisance parameter is not identified under the null hypothesis. *Rochester Center for Economic Research Working Paper No. 296*, 1991.
- B.E. Hansen. The likelihood ratio test under nonstandard conditions: testing the markov switching model of gnp. *Journal of applied Econometrics*, 7(S1), 1992.
- B.E. Hansen. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica: Journal of the econometric society*, pages 413–430, 1996.
- A.M. Hasofer. Upcrossings of random fields. *Advances in Applied Probability*, 10: 14–21, 1978.
- Y. Hochberg. A sharper bonferroni procedure for multiple tests of significance. *Biometrika*, 75(4):800–802, 1988.
- H. Hotelling. Tubes and spheres in n-spaces, and a class of statistical problems. *American Journal of Mathematics*, 61(2):440–460, 1939.
- J. Hüsler. Asymptotic approximation of crossing probabilities of random sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 63(2):257–270, 1983.
- J. Hüsler. Extreme values of non-stationary random sequences. *Journal of applied probability*, 23(4):937–950, 1986.

- G. James et al. *An introduction to statistical learning*, volume 112. Springer, 2013.
- H.C. Johnston. Cliques of a graph-variations on the bron-kerbosch algorithm. *International Journal of Computer & Information Sciences*, 5(3):209–238, 1976.
- B.J. Kavanagh. Overview: Dark matter. *DM-Stat workshop: Statistical Challenges in the Search for Dark Matter, Banff, Canada*, 2018.
- M. Knowles and D. Siegmund. On hotelling’s approach to testing for a nonlinear parameter in regression. *International Statistical Review/Revue Internationale de Statistique*, pages 205–220, 1989.
- P. Koppenburg, Z. Dolezal, and M. Smizanska. Rare decays of b hadrons. *arXiv:1606.00999*, 2016.
- R.O. Kuehl. *Statistical principles of research design and analysis*. Duxbury Resource Center, 1994.
- M.R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and related properties of random sequences and processes*. Springer-Verlag New York Inc., 1983.
- J.T. Leek and R.D. Peng. Statistics: P values are just the tip of the iceberg. *Nature*, 520(7549):612, 2015.
- Y. Lin and B.G. Lindsay. Projections on cones, chi-bar squared distributions, and weyl’s formula. *Statistics & probability letters*, 32(4):367–376, 1997.
- G. Lindgren. Extreme values and crossings for the χ^2 -process and other functions of multidimensional gaussian processes, by reliability applications. *Advances in Applied Probability*, 12(3):746–774, 1980.
- L. Lyons. Discovering the significance of 5 sigma. *arXiv preprint arXiv:1310.1284*, 2013.
- P.A.P. Moran. On asymptotically optimal tests of composite hypotheses. *Biometrika*, pages 47–55, 1970.
- V.M.R. Muggeo et al. Segmented: an r package to fit regression models with broken-line relationships. *R news*, 8(1):20–25, 2008.
- J.H. Oort. The force exerted by the stellar system in the direction perpendicular to the galactic plane and some related problems. *Bulletin of the Astronomical Institutes of the Netherlands*, 6:249, 1932.

- R.S. Pilla and C. Loader. Inference in perturbation models, finite mixtures and scan statistics: the volume-of-tube formula. *arXiv preprint arXiv:math/0511503*, 2005.
- R.S. Pilla, C. Loader, and C.C. Taylor. New technique for finding needles in haystacks: Geometric approach to distinguishing between a new source and random fluctuations. *Physical review letters*, 95(23):230202, 2005.
- R.E. Quandt. A comparison of methods for testing nonnested hypotheses. *The Review of Economics and Statistics*, pages 92–99, 1974.
- M. Raab. *On the number of exceedances in Gaussian and related sequences*. PhD thesis, 1997.
- S.O. Rice. Mathematical analysis of random noise. *Bell Labs Technical Journal*, 23(3):282–332, 1944.
- R.J. Serfling. *Approximation theorems of mathematical statistics*, volume 162. John Wiley & Sons, 2009.
- K. Sharpe. Some properties of the crossings process generated by a stationary χ^2 -process. *Advances in Applied Probability*, 10(2):373–391, 1978.
- J. Sun. Tail probabilities of the maxima of gaussian random fields. *The Annals of Probability*, pages 34–71, 1993.
- A. Takemura and S. Kuriki. Weights of $\bar{\chi}^2$ distribution for smooth or piecewise smooth cone alternatives. *The Annals of Statistics*, pages 2368–2387, 1997.
- A. Takemura and S. Kuriki. On the equivalence of the tube and euler characteristic methods for the distribution of the maximum of gaussian fields over piecewise smooth domains. *Annals of Applied Probability*, pages 768–796, 2002.
- J.E. Taylor and R.J. Adler. Euler characteristics for gaussian fields on manifolds. *Annals of Probability*, pages 533–563, 2003.
- J.E. Taylor and K.J. Worsley. Random fields of multivariate test statistics, with applications to shape analysis. *The Annals of Statistics*, pages 1–27, 2008.
- J.E. Taylor and K.J. Worsley. Detecting sparse cone alternatives for gaussian random fields, with an application to fmri. *Statistica Sinica*, pages 1629–1656, 2013.
- J.E. Taylor et al. Validity of the expected euler characteristic heuristic. *The Annals of Probability*, 33(4):1362–1396, 2005.

- T.M. Undagoitia and L. Rauch. Dark matter direct-detection experiments. *Journal of Physics G: Nuclear and Particle Physics*, 43(1):013001, 2015.
- D.A. van Dyk. The role of statistics in the discovery of a higgs boson. *Annual Review of Statistics and Its Application*, 1:41–59, 2014.
- E. Vanmarcke. *Random fields: analysis and synthesis*. World scientific, 2010.
- O. Vitells and E. Gross. Estimating the significance of a signal in a multi-dimensional search. *Astroparticle Physics*, 35(5):230–234, 2011.
- R.L. Wasserstein and N.A. Lazar. The asa’s statement on p-values: context, process, and purpose. 2016.
- H. Weyl. On the volume of tubes. *American Journal of Mathematics*, 61(2):461–472, 1939.
- S.S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62, 1938.
- K.J. Worsley. Local maxima and the expected euler characteristic of excursion sets of χ^2 , f and t fields. *Advances in Applied Probability*, 26(1):13–42, 1994.
- K.J. Worsley. Estimating the number of peaks in a random field using the hadwiger characteristic of excursion sets, with applications to medical images. *The Annals of Statistics*, pages 640–669, 1995.
- F. Zwicky. Spectral displacement of extra galactic nebulae. *Helvetica Physica Acta*, 6:110, 1933.
- F. Zwicky. On the masses of nebulae and of clusters of nebulae. *The Astrophysical Journal*, 86:217, 1937.



RATES OF CONVERGENCE AND GOODNESS OF THE APPROXIMATIONS

Most of the results discussed in Chapters 3 and 4 focus on bounding/approximating excursion probabilities of the form

$$P\left(\sup_{\theta \in \Theta} \{W(\theta)\} > c\right) \tag{A.1}$$

where $\{W(\theta)\}$ in (A.1) is a generic stochastic process indexed by $\theta \in \Theta$. In this appendix, we limit our attention to the case where $\{W(\theta)\}$ is either a Gaussian, a χ^2 or a $\bar{\chi}_{01}^2$ process, and we quantify the rate at which the right hand sides (1.13), (2.4) and (3.10) converge to their left hand sides. These approximations are also validated via a suite of numerical studies throughout Chapters 2, 3 and 4.

A.1 RATES OF CONVERGENCE

We would like some indication on the sharpness of the bound in (3.6), i.e.,

$$P\left(\sup_{\theta \in \Theta} \{W(\theta)\} > c\right) \leq P(W(\mathfrak{L}) > c) + \frac{a(c)}{a(c_0)} E[N_{c_0}] \quad \forall c_0 \leq c, c_0 \in \mathbb{R}, \tag{A.2}$$

for the normal, χ_s^2 and $\bar{\chi}_{01}^2$ cases. As discussed in Chapters 1, 3 and 4 under Condition 1.1.4 and Conditions 3.1.1, the number of upcrossings N_c converges to a

Poisson process as $c \rightarrow \infty$ (e.g., Falk et al. [2010, p. 364]). Thus, assuming that $E[N_c]$ converges to a finite limit μ , we write

$$P(N_c \geq 1) \rightarrow 1 - e^{-\mu}. \quad (\text{A.3})$$

In this section, we focus on $\{Z(\theta)\}$ and $\{T(\theta)\}$, i.e., the Gaussian and χ_s^2 processes defined in Chapter 1. Results pertaining to χ_s^2 processes naturally extend to $\bar{\chi}_{01}^2$ processes since the process of upcrossings is governed by its χ_1^2 component for all $c > 0$.

Condition 1.1.4 and Conditions 3.1.1 can be formalized in more details considering the covariance function $\rho(\theta, \theta^\dagger)$ of the underlying process [e.g., Pickands, 1969a, Lindgren, 1974, 1980b, Aronowich and Adler, 1985, Tan and Hashorva, 2013, Liu and Ji, 2014, Hashorva and Ji, 2015]. Specifically, for non-stationary processes, we exploit the results of Tan and Hashorva [2013], Liu and Ji [2014], and Hashorva and Ji [2015], and we require (A.4), and (A.5) to be satisfied for some $p, q \in (0, 2]$ and some positive constants A, B

$$\rho(\theta, \theta^\dagger) = 1 - A|\theta - \theta^\dagger|^p + o(|\theta - \theta^\dagger|^p), \quad \text{as } |\theta - \theta^\dagger| \rightarrow 0 \quad (\text{A.4})$$

$$\rho(\theta, \mathfrak{U}) = 1 - B|\theta - \mathfrak{U}|^q + o(|\theta - \mathfrak{U}|^q) \quad \text{as } |\theta - \mathfrak{U}| \rightarrow 0 \quad (\text{A.5})$$

where \mathfrak{U} is the upper bound of the parameter space of θ , i.e., $\Theta \in [\mathfrak{L}, \mathfrak{U}]$. It follows that (3.10) and (1.13) can be rewritten as in (A.6) and (A.7), respectively

$$P\left(\sup_{\theta \in \Theta} \{Z(\theta)\} > c\right) = e^{-\frac{c^2 - c_0^2}{2}} E[N_{c_0}^Z] + o(c^{\max(\frac{2}{p} - \frac{2}{q}, 0) - 1} e^{-c^2/2}) \quad (\text{A.6})$$

$$P\left(\sup_{\theta \in \Theta} \{T(\theta)\} > c\right) = \left(\frac{c}{c_0}\right)^{\frac{s-1}{2}} e^{-\frac{c-c_0}{2}} E[N_{c_0}^T] + o(c^{\max(\frac{2}{p} - \frac{1}{q}, 0) + s/2 - 1} e^{-c/2}). \quad (\text{A.7})$$

Notice that, as $c \rightarrow \infty$, the first terms on the right hand sides of (1.13) and (3.10) are dominated by the second term, thus the second terms in the right hand sides of (A.6) and (A.7) also account for this.

For the stationary case, following Lindgren [1980a,b], (A.6) and (A.7) simplify to (A.8) and (A.9) respectively

$$P\left(\sup_{\theta \in \Theta} \{Z(\theta)\} > c\right) = e^{-\frac{c^2 - c_0^2}{2}} E[N_{c_0}^Z] + o(ce^{-c^2/2}) \quad (\text{A.8})$$

$$P\left(\sup_{\theta \in \Theta} \{T(\theta)\} > c\right) = \left(\frac{c}{c_0}\right)^{\frac{s-1}{2}} e^{-\frac{c-c_0}{2}} E[N_{c_0}^T] + o(c^{s/2-1} e^{-c/2}). \quad (\text{A.9})$$

The error rates in (A.6) and (A.7) refer to the asymptotic expansions provided by Tan and Hashorva [2013], Liu and Ji [2014], Hashorva and Ji [2015]. They approximate the excursion probabilities on the left hand sides of (A.6) and (A.7) via the so-called *geometric approach* of Piterbarg, and based on Minkowski functionals, i.e., the Euclidean version of the Lipschitz-Killing curvatures introduced in Chapter 6 (the reader is directed to Pickands [1969a,b], Adler [2000], Adler and Taylor [2009] and Piterbarg [2012] for further details). Thus, the expansions obtained via the geometric approach do not consider directly the expected number of upcrossings. However, as noted in Adler and Taylor [2009, p. 386], the approach pioneered by Piterbarg is indirectly leading to the expected Euler characteristic and thus, in one dimension, to the expected number of upcrossings. Consequently, we expect the error rates of the two approaches to coincide and (A.6) and (A.7) to hold.

A.2 APPROXIMATING TOHM VIA BONFERRONI

Results 4.1.5 and 4.1.6 formalize the Poisson convergence of the processes of exceedances and upcrossings of c by $\{W(\theta_r)\}$, which in turn guarantee the validity of Theorem 4.2.1. Thus, one way to assess the goodness of the approximation in (4.13) is to measure the distance between the distribution of the process of exceedances, \dot{N}_{c_R} , with intensity p_{BF} (see 4.8), and the Poisson process with intensity $E[\tilde{N}_{c_R}]$, namely $\mathcal{P}(E[\tilde{N}_{c_R}])$.

In order to achieve this goal, we consider the variational distance $d(\cdot, \cdot)$ between \dot{N}_{c_R} and the Poisson process $\mathcal{P}(E[\tilde{N}_{c_R}])$ and we let $\mathcal{P}(p_{BF})$ be the Poisson process with mean p_{BF} . For the normal sequence $\{Z(\theta_r)\}$, it follows from Barbour and Holst [1989] (see also Falk et al. [2010, p. 372]) that, under the conditions required by Theorem 4.2.1, $d(\dot{N}_{c_R}^Z, \mathcal{P}(p_{BF}))$ specifies

$$d\left(\dot{N}_{c_R}^Z, \mathcal{P}(p_{BF})\right) \leq \frac{1 - e^{p_{BF}}}{p_{BF}} \left(\frac{p_{BF}^2}{R} + \sum_{r \neq r' \leq R} \left| \text{cov}(\mathbb{1}_{\{Z(\theta_r) > c_R\}}, \mathbb{1}_{\{Z(\theta_{r'}) > c_R\}}) \right| \right). \quad (\text{A.10})$$

Thus, using the triangular inequality property of the variational distance we can

write $d(\dot{N}_{c_R}^Z, \mathcal{P}(E[\tilde{N}_{c_R}^Z]))$ as

$$\begin{aligned}
d\left(\dot{N}_{c_R}^Z, \mathcal{P}(E[\tilde{N}_{c_R}^Z])\right) &\leq d\left(\dot{N}_{c_R}^Z, \mathcal{P}(p_{BF})\right) + d\left(\mathcal{P}(p_{BF}), \mathcal{P}(E[\tilde{N}_{c_R}^Z])\right) \\
&= \frac{1 - e^{p_{BF}}}{p_{BF}} \left(\frac{p_{BF}^2}{R} + \sum_{r \neq r' \leq R} \left| \text{COV}(\mathbb{1}_{\{Z(\theta_r) > c_R\}}, \mathbb{1}_{\{Z(\theta_{r'}) > c_R\}}) \right| \right) \\
&\quad + \left| E[\tilde{N}_{c_R}^Z] - p_{BF} \right|.
\end{aligned} \tag{A.11}$$

More work has to be done for the χ_s^2 case. Specifically let $\{T(\theta_r)\}$ be a χ_s^2 sequence where each component is the sum of the squares of the standard normal sequences $\{Z_1(\theta_r)\}, \dots, \{Z_s(\theta_r)\}$. Following the approach of Raab [1997, Sec. 2.3.2], we consider the indicator functions

$$\mathbb{1}_r = \mathbb{1}_{\{h(\pm Z_1(\theta_r)^2, \dots, \pm Z_s(\theta_r)^2) > c_R\}} \tag{A.12}$$

where $h(z_1, \dots, z_s) = (z_1^2 + \dots + z_s^2) \mathbb{1}_{\{z_1 \geq 0, \dots, z_s \geq 0\}}$. We can specify the variational distance between $\dot{N}_{c_R}^T$ and $\mathcal{P}(p_{BF})$ as

$$d\left(\dot{N}_{c_R}^T, \mathcal{P}(p_{BF})\right) \leq \frac{1 - e^{p_{BF}}}{p_{BF}} \left(\frac{p_{BF}^2}{R} + \sum_{r=1}^R \sum_{r' \neq r}^R |\text{COV}(\mathbb{1}_r, \mathbb{1}_{r'})| \right). \tag{A.13}$$

As in (A.11), it follows that

$$\begin{aligned}
d\left(\dot{N}_{c_R}^T, \mathcal{P}(E[\tilde{N}_{c_R}^T])\right) &\leq d\left(\dot{N}_{c_R}^T, \mathcal{P}(p_{BF})\right) + d\left(\mathcal{P}(p_{BF}), \mathcal{P}(E[\tilde{N}_{c_R}^T])\right) \\
&= \frac{1 - e^{p_{BF}}}{p_{BF}} \left(\frac{p_{BF}^2}{R} + \sum_{r=1}^R \sum_{r' \neq r}^R |\text{COV}(\mathbb{1}_r, \mathbb{1}_{r'})| \right) \\
&\quad + \left| E[\tilde{N}_{c_R}^T] - p_{BF} \right|.
\end{aligned} \tag{A.14}$$

Since in the $\bar{\chi}_{01}^2$ case, the process of upcrossings depends only on the χ_1^2 component, the bound in (A.14) also applies to the $\bar{\chi}_{01}^2$.

References in Appendix

- R.J. Adler. On excursion sets, tube formulas and maxima of random fields. *Annals of Applied Probability*, pages 1–74, 2000.
- R.J. Adler and J.E. Taylor. *Random fields and geometry*. Springer Science & Business Media, 2009.
- M. Aronowich and R.J. Adler. Behaviour of χ^2 processes at extrema. *Advances in applied probability*, 17(2):280–297, 1985.
- A.D. Barbour and L. Holst. Some applications of the stein-chen method for proving poisson convergence. *Advances in Applied Probability*, 21(1):74–90, 1989.
- M. Falk, J. Hüsler, and R.-D. Reiss. *Laws of small numbers: extremes and rare events*. Springer Science & Business Media, 2010.
- E. Hashorva and L. Ji. Piterbarg theorems for chi-processes with trend. *Extremes*, 18(1):37–64, 2015.
- G. Lindgren. A note on the asymptotic independence of high level crossings for dependent gaussian processes. *The Annals of Probability*, pages 535–539, 1974.
- G. Lindgren. Point processes of exits by bivariate gaussian processes and extremal theory for the χ^2 process and its concomitants. *Journal of Multivariate Analysis*, 10(2):181–206, 1980a.
- G. Lindgren. Extreme values and crossings for the x^2 -process and other functions of multidimensional gaussian processes, by reliability applications. *Advances in Applied Probability*, 12(3):746–774, 1980b.
- P. Liu and L. Ji. Extremes of chi-square processes with trend. *arXiv preprint arXiv:1407.6501*, 2014.
- J. Pickands. Upcrossing probabilities for stationary gaussian processes. *Transactions of the American Mathematical Society*, 145:51–73, 1969a.
- J. Pickands. Asymptotic properties of the maximum in a stationary gaussian process. *Transactions of the American Mathematical Society*, 145:75–86, 1969b.

- V.I. Piterbarg. *Asymptotic methods in the theory of Gaussian processes and fields*, volume 148. American Mathematical Soc., 2012.
- M. Raab. *On the number of exceedances in Gaussian and related sequences*. PhD thesis, 1997.
- Z. Tan and E. Hashorva. Exact asymptotics and limit theorems for supremum of stationary χ -processes over a random interval. *Stochastic Processes and their Applications*, 123(8):2983–2998, 2013.