Estimating the "look elsewhere effect" when searching for a signal

Ofer Vitells
Weizmann Institute of Science, Rehovot 76100, Israel

Abstract
The "look elsewhere effect" refers to a common situation where one searches for a signal in some space of parameters - for example, a resonance search with unknown mass, or a search for astrophysical point sources with unknown location in the sky. Since Wilks' theorem does not apply in such cases, one usually has to resort to computationally expansive Monte-Carlo simulations in order to correctly estimate the significance of a given observation. Recent results from the theory of random fields provide powerful tools which may be used to alleviate this difficulty, in a wide range of applications. We review those results and discuss their implementation in problems of practical interest.

1 Introduction

Experiments that aim at discovering new physical phenomena often involve a search for a signal over some space of continuous parameters. One such example is the search for the Higgs boson at particle colliders, where one searches for a peak within some range of an invariant mass distribution. Another example is the search for astrophysical neutrino sources that can be located at any direction in the sky. To assess the significance of a local deviation from the background hypothesis in terms of a p-value, one needs to take into account the probability of such a deviation to occur anywhere within the search range. This is the so called "look elsewhere effect". Estimation of the p-value could be performed by repeated Monte Carlo simulations of the experiment's outcome under the background-only hypothesis, but this approach could be highly time consuming since for each of those simulations the entire search procedure needs to be applied to the data, and to establish a discovery claim at the 5σ level (p-value=2.87×10^{-7}) the simulation needs to be repeated at least $\mathcal{O}(10^7)$ times. Fortunately, recent advances in the theory of random fields provide analytical tools that can be used to address exactly such problems, in a wide range of experimental settings. Such methods could be highly valuable for experiments searching for signals over large parameter spaces, as the reduction in necessary computation time can be dramatic. Random field theoretic methods were first applied to the statistical hypothesis testing problem in [1], for some special case of a one dimensional problem. A practical implementation of this result, aimed at the high-energy physics community, was made in [2]. Similar results for some cases of multi-dimensional problems [3] [4] were applied to statistical tests in the context of brain imaging [5]. More recently, a generalized result dealing with random fields over arbitrary Riemannian manifolds was obtained [6], openning the door for a plethora of new possible applications. Here we discuss the implementation of these results in the context of physics experiments, taking two representative cases as specific examples. In section 2 the general framework of an hypothesis test is briefly presented with connection to random fields. In section 3 the main theoretical result is presented, and two examples are treated in detail in sections 4 and 5.

2 Formalism of a search as a statistical test

Consider the problem of an hypothesis testing, where one tests the background (null) hypothesis $H_0 : \mu = 0$, against a signal hypothesis $H_1 : \mu > 0$, where μ represents the signal strength. Suppose that θ are some nuisance parameters describing other properties of the signal (such as location), which are therefore not present under the null. Additional nuisance parameters, denoted by θ', may be present under both hypotheses. Denote by $\mathscr{L}(\mu, \theta, \theta')$ the likelihood function. One may then construct the profile likelihood ratio test statistic [7]

$$q = -2\log \frac{\max_{\theta'} \mathscr{L}(\mu = 0, \theta')}{\max_{\mu,\theta,\theta'} \mathscr{L}(\mu, \theta, \theta')} \tag{1}$$

and reject the null hypothesis if the test statistic is larger then some critical value. Note that when the signal strength is set to zero the likelihood by definition does not depend on θ, and the test statistic (1) can therefore be written as

$$q = \max_{\theta \in \mathscr{M}} q(\theta) \tag{2}$$

where $q(\theta)$ is the profile likelihood ratio with the signal nuisance parameters fixed to the point θ, and we have denoted by $\mathscr{M}$ the D-dimensional manifold to which the parameters θ belong. Under the conditions of Wilks' theorem [8], $q(\theta)$ follows a χ^2 distribution with one degree of freedom when the null hypothesis is true. When viewed as a function over the manifold $\mathscr{M}$, $q(\theta)$ is therefore a χ^2 *random field*, namely a set of random variables that are continuously mapped to the manifold $\mathscr{M}$. To quantify the significance of a given observation in terms of a p-value, one is required to calculate the probability of the maximum of the field to be above some level, that is, the excursion probability of the field:

$$p\text{-value} = \mathbb{P}[\max_{\theta \in \mathscr{M}} q(\theta) > u]. \tag{3}$$

In most cases, direct calculation of the above quantity will probably be too difficult to be of any practical use. However, other closely related quantities exist for which surprisingly simple closed-form expressions have been derived under general conditions. Those will allow to estimate the excursion probability (3) when the level u is large, which is the main region of interest.

3 The excursion sets of random fields

The excursion set of a field above a level u, denoted by A_u, is defined as the set of points θ for which the value of the field $q(\theta)$ is larger than u,

$$A_u = \{\theta \in \mathscr{M} : q(\theta) > u\} \tag{4}$$

and we will denote by $\phi(A_u)$ the Euler characteristic of the excursion set A_u. A fundamental result of [6] states that the expectation of the Euler characteristic $\phi(A_u)$ is given by the following expression:

$$\mathbb{E}[\phi(A_u)] = \sum_{d=0}^{D} \mathscr{N}_d \rho_d(u). \tag{5}$$

The coefficients $\mathscr{N}_d$ are related to some geometrical properties of the manifold and the covariance structure of the field, for our purposes however they will be just a set of unknown constants. The functions $\rho_d(u)$ are 'universal' in the sense that they are determined only by the distribution type of the field $q(\theta)$, and their analytic expressions are known for a large class of 'Gaussian related' fields, such as χ^2 with arbitrary degrees of freedom. The zeroth order term of eq. (5) is a special case for which $\mathscr{N}_0$ and $\rho_0(u)$ are generally given by

$$\mathscr{N}_0 = \phi(\mathscr{M}), \qquad \rho_0(u) = \mathbb{P}[q(\theta) > u] \tag{6}$$

Namely, $\mathscr{N}_0$ is the Euler characteristic of the entire manifold and $\rho_0(u)$ is the tail probability of the distribution of the field. (Note that when the manifold is reduced to a point, this result becomes trivial).

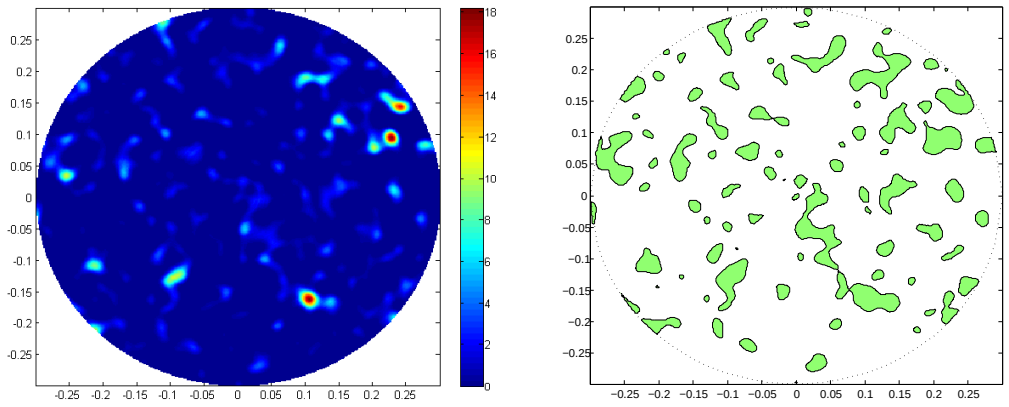
When the level u is high enough, excursions above u become rare and the excursion set becomes a few disconnected hyper-ellipses. In that case the Euler characteristic $\phi(A_u)$ simply counts the number of disconnected components that make up A_u. For even higher levels this number is mostly zero and rarely one, and Its expectation therefore converges asymptotically to the excursion probability. We can thus use it as an approximation to the excursion probability for large enough u [9]

$$\mathbb{E}[\phi(A_u)] \approx \mathbb{P}[\max_{\theta \in \mathscr{M}} q(\theta) > u]. \tag{7}$$

The practical importance of Eq. (5) now becomes clear, as it allows to estimate the excursion probabilities above high levels. Furthermore, the problem is reduced to finding the constants $\mathscr{N}_d, d > 0$. Since Eq. (5) holds for any level u, this could be achieved simply by calculating the average of $\phi(A_u)$ at some low levels, which can be done using a small set of Monte Carlo simulations. We shall now turn to a few examples where this procedure is demonstrated.

4 Example 1: detecting neutrino sources

The IceCube experiment [10] is a neutrino telescope located at the south pole and aimed at detecting astrophysical neutrino sources. The detector measures the energy and angular direction of incoming neutrinos, trying to distinguish an astrophysical point-like signal from a large background of atmospheric neutrinos spread across the sky. The nuisance parameters over which the search is performed are therefore the angular coordinates (θ, φ)[1]. We follow [11] for the definitions of the signal and background distributions and calculate a profile likelihood ratio as described in the previous section. Figure 1 shows a "significance map" of the sky, namely the values of the test statistic $q(\theta, \varphi)$ as well as the corresponding excursion set above $q = 1$. To reduce computation time we restrict here the search space to the portion of the sky at declination angle 27° below the zenith, however all the features of a full sky search are maintained. Note that the most significance point has a value of the test statistic above 16, which would correspond to a significance exceeding 4σ if this point would have been analyzed alone, that is without the "look elsewhere" effect.



(a) (b)

Fig. 1: (a) A significance map showing the test statistic $q(\theta, \varphi)$ for a background simulation (b) The corresponding excursion set above $q = 1$.

[1]The signal model may include additional parameters such as spectral index and time, which we do not consider here for simplicity.

For a χ^2 random field with one degree of freedom and for two search dimensions, Eq. (5) reads [6]

$$\mathbb{E}[\phi(A_u)] = \mathbb{P}[\chi^2 > u] + e^{-u/2}(\mathcal{N}_1 + \sqrt{u}\mathcal{N}_2). \tag{8}$$

To estimate the the coefficients $\mathcal{N}_1, \mathcal{N}_2$ we use a set of 20 background simulations, and calculate the Euler characteristic of the excursion set corresponding to the levels $u = 0, 1$. This gives the estimates $\mathbb{E}[\phi(A_0)] = 33.5 \pm 2$ and $\mathbb{E}[\phi(A_1)] = 94.6 \pm 1.3$. By solving for the unknown coefficients we obtain $\mathcal{N}_1 = 33 \pm 2$ and $\mathcal{N}_2 = 123 \pm 3$. The prediction of Eq. (8) is then compared against a set of approx. 200,000 background simulations, where for each one the maximum of $q(\theta, \varphi)$ is found by scanning the entire range. The results are shown in Figure 2. As expected, the approximation becomes better as the p-value becomes smaller. The agreement between Eq. (8) and the observed p-value is maintained up to the smallest p-value that the available statistics allows us to estimate.

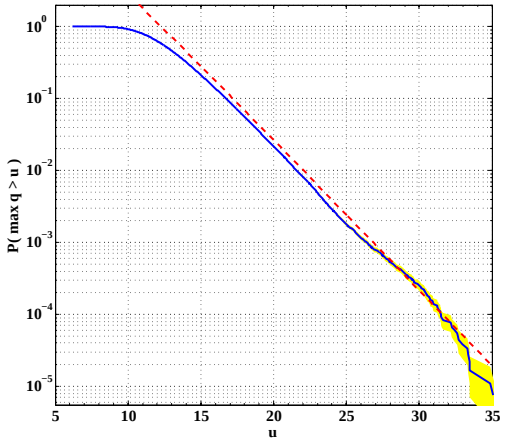


Fig. 2: The prediction of Eq. (8) (dashed red) against the observed p-value (solid blue) from a set of 200,000 background simulations. The yellow band represents the statistical uncertainty due to the available number of background simulations.

4.1 Slicing the parameter space

A useful property of Eq. (5) that can be illustrated by this example, is the ability to consider only a small 'slice' of the parameter space from which the expected Euler characteristic (and hence p-value) of the entire space can be estimated, if a symmetry is present in the problem. This can be done using the 'inclusion-exclusion' property of the Euler characteristic:

$$\phi(A \cup B) = \phi(A) + \phi(B) - \phi(A \cap B). \tag{9}$$

Since the neutrino background distribution is assumed to be uniform in azimuthal angle (φ), we can divide the sky to N identical slices of azimuthal angle, as illustrated in Figure 3. Applying (9) to this case, the expected Euler characteristic is given by

$$\mathbb{E}[\phi(A_u)] = N \times (\mathbb{E}[\phi(slice)] - \mathbb{E}[\phi(edge)]) + \mathbb{E}[\phi(0)] \tag{10}$$

where an 'edge' is the line common to two adjacent slices, and $\phi(0)$ is the Euler characteristic of the point at the origin (see Figure 3).

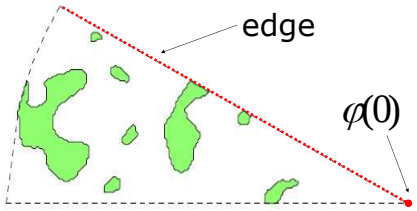


Fig. 3: Illustration of the excursion set in a slice of a sky, showing also an edge (dashed red) and the origin as defined in Eq. (10).

We can now apply Eq. (5) to both $\phi(slice)$ and $\phi(edge)$ and estimate the corresponding coefficients as was done before, using only simulations of a single slice of the sky. Following this procedure we obtain for this example with $N = 18$ slices from 40 background simulations, $\mathscr{N}_1^{slice} = 6 \pm 0.5, \mathscr{N}_2^{slice} = 6.7 \pm 0.8$ and $\mathscr{N}_1^{edge} = 4.4 \pm 0.2$. Using (10) this leads to the full sky coefficients $\mathscr{N}_1 = 28 \pm 9$ and $\mathscr{N}_2 = 120 \pm 14$, a result which is consistent with the full sky simulation procedure.

5 Example 2: A 'mass bump' with unknown width

As a second example we consider the common problem in high energy physics of detecting a 'mass bump' on top of a continuous background, and we assume that the width of the bump is also a-priori unknown (within some range). We assume an exponential background distribution and a gaussian signal distribution, with the signal location and width being the free nuisance parameters. The search space is therefore two dimensional in this problem as well. Figure 4 shows an example histogram of background events with a signal best fit. Figure 5 shows the values of the test statistic q as a function of the mass and the width, and the corresponding excursion set above $q = 1$.

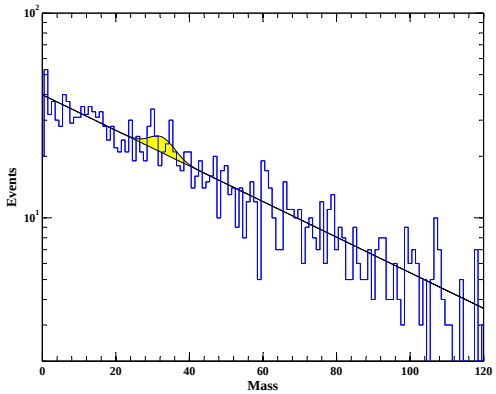
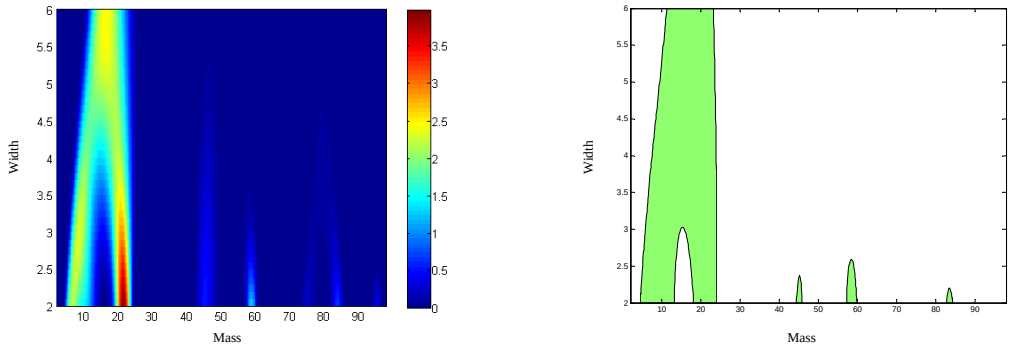


Fig. 4: An example histogram of background events with the signal best fit shown in yellow.

The search space (mass,width) has clearly a non trivial correlation structure, which is evident in Figure 5. The procedure for estimating the p-value is nevertheless identical to that of the previous



(a) (b)

Fig. 5: (a) The values of the test statistic q as a function of the signal mass and the width (b) The corresponding excursion set above $q = 1$.

example, and the expected Euler characteristic is similarly given by Eq. (8), where the difference is only in the numerical values of the coefficients $\mathscr{N}_1, \mathscr{N}_2$. Here we find $\mathscr{N}_1 = 4 \pm 0.2$ and $\mathscr{N}_2 = 0.7 \pm 0.3$, and the predicted p-value is shown in Figure 6 compared to the observed p-value from a set of 200,000 background simulations. Again we find an excellent agreement between the Euler characteristic formula and the observed p-value, demonstrating the usefulness of this result.

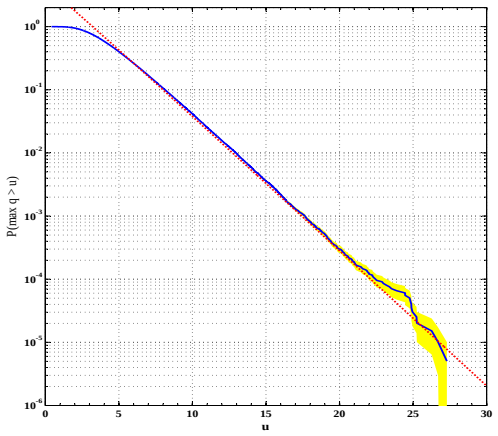


Fig. 6: The prediction of Eq. (8) (dashed red) against the observed p-value (solid blue) from a set of 200,000 background simulations. The yellow band represents the statistical uncertainty due to the available number of background simulations.

6 Summary

The Euler characteristic formula, a fundamental result from the theory of random fields, provides a practical mean of estimating a p-value while taking into account the "look elsewhere effect". This result is valid under general conditions and is therefore applicable to a wide range of problems, as we have demonstrated in the two representative cases studied in this work. This could greatly ease the compu-

tational burden of having to perform large number of Monte Carlo simulations, required to establish a discovery claim.

Acknowledgements

We are grateful to Michael Woodroofe and Luc Demortier, for their helpful comments and discussions during the 2010 Banff workshop on statistical issues [12]. We thank Jim Braun and Teresa Montaruli for their help in providing us the background simulation data of IceCube which was used to perform this analysis.

References

[1] R.B. Davies, *Hypothesis testing when a nuisance parameter is present only under the alternative.*, Biometrika **74** (1987), 33-43.

[2] E. Gross and O. Vitells, *Trial factors for the look elsewhere effect in high energy physics* , Eur. Phys. J. C, **70** (2010), 525-530.

[3] R.J. Adler and A.M. Hasofer, *Level Crossings for Random Fields*,Ann. Probab. **4**, Number 1 (1976), 1-12.

[4] R.J. Adler, *The Geometry of Random Fields*, New York (1981), Wiley, ISBN: 0471278440.

[5] K.J. Worsley, S. Marrett, P. Neelin, A.C. Vandal, K.J. Friston and A.C. Evans, *A Unified Statistical Approach for Determining Significant Signals in Location and Scale Space Images of Cerebral Activation*, Human Brain Mapping **4** (1996) 58-73.

[6] R.J. Adler and J.E. Taylor, *Random Fields and Geometry* , Springer Monographs in Mathematics (2007). ISBN: 978-0-387-48112-8.

[7] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C **71** (2011) 1544, [arXiv:1007.1727].

[8] S.S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, Ann. Math. Statist. **9** (1938) 60-62.

[9] J. Taylor, A. Takemura and R.J. Adler, *Validity of the expected Euler characteristic heuristic*, Ann. Probab. 33 (2005) 1362-1396.

[10] J. Ahrens et al. and The IceCube Collaboration, Astropart. Phys. **20** (2004), 507.

[11] J. Braun, J. Dumma, F. De Palmaa, C. Finleya, A. Karlea and T. Montaruli, *Methods for point source analysis in high energy neutrino telescopes*, Astropart. Phys. **29** (2008) 299-305 [arXiv:0801.1604].

[12] Banff International Research Station Workshop on Statistical issues relevant to significance of discovery claims, http://www.birs.ca/events/2010/5-day-workshops/10w5068.