

# A method for comparing non-nested models with application to astrophysical searches for new physics

Sara Algeri,<sup>1,2★</sup> Jan Conrad<sup>1,2,3†</sup> and David A. van Dyk<sup>1</sup>

<sup>1</sup>Statistics Section, Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, UK

<sup>2</sup>Department of Physics, Stockholm University, AlbaNova, SE-106 91 Stockholm, Sweden

<sup>3</sup>The Oskar Klein Centre for Cosmoparticle Physics, AlbaNova, SE-106 91 Stockholm, Sweden

Accepted 2016 February 10. Received 2016 February 10; in original form 2015 December 25

## ABSTRACT

Searches for unknown physics and decisions between competing astrophysical models to explain data both rely on statistical hypothesis testing. The usual approach in searches for new physical phenomena is based on the statistical likelihood ratio test and its asymptotic properties. In the common situation, when neither of the two models under comparison is a special case of the other i.e. when the hypotheses are non-nested, this test is not applicable. In astrophysics, this problem occurs when two models that reside in different parameter spaces are to be compared. An important example is the recently reported excess emission in astrophysical  $\gamma$ -rays and the question whether its origin is known astrophysics or dark matter. We develop and study a new, simple, generally applicable, frequentist method and validate its statistical properties using a suite of simulations studies. We exemplify it on realistic simulated data of the *Fermi*-Large Area Telescope  $\gamma$ -ray satellite, where non-nested hypotheses testing appears in the search for particle dark matter.

**Key words:** astroparticle physics – methods: data analysis – methods: statistical – dark matter.

## 1 MODEL COMPARISON IN ASTROPARTICLE PHYSICS

In astrophysics, hypothesis testing is ubiquitous, because progress is made by comparing competing models to experimental data. In the special case, where new physical phenomena are searched for, the most common choice of hypothesis test is the likelihood ratio test (LRT), whose popularity is partly motivated by the fact that, assuming  $H_0$  is true, the asymptotic distribution of the LRT statistic is a  $\chi^2$ . Such result holds if the regularity conditions specified in Wilks's theorem hold (Wilks 1938). A key necessary condition is ‘nested-ness’, meaning that there is a full model of which both the models under  $H_0$  and the alternative hypothesis,  $H_1$ , are special cases. This is obviously the case for the search for new particles where the null hypothesis (or baseline model),  $H_0$ , is given by ‘background’ and  $H_1$  is given by ‘background+signal of new particle’. However, cases where model comparison is non-nested are common: for instance, when a known astrophysical signal can be confused with new physics, see Ackermann et al. (2012) for an example from astroparticle physics, or if the models to be compared reside in different parameter spaces (Profumo & Linden 2012); as in  $\gamma$ -ray bursts (Guiriec et al. 2015). In these situations, Monte Carlo

simulations of the measurement process are often the only possibility, but are challenged by stringent significance requirements, e.g. at the  $5\sigma$  level. We present a solution that allows evaluation of accurate statistical significances for non-nested model comparison while avoiding extensive Monte Carlo simulations. As a concrete example, we apply the proposed procedure to the search for particle dark matter, where the method has particular importance.

One way to search for dark matter is to consider its hypothesized annihilation products, i.e.  $\gamma$ -rays, that can be detected by space-borne or ground-based  $\gamma$ -rays telescopes (Conrad 2005; Cohen-Tanugi & Stigari 2015). Here, the issue of source confusion is one of the most challenging aspects of claiming discovery of a dark matter induced signal. A detected excess of  $\gamma$ -rays may either originate from dark matter annihilation or be caused by conventional, known astrophysical sources. Discrimination can be performed using their spectral distributions, however these are not necessarily part of the same parameter space (see below). This situation arises for example in the search for dark matter sources among the unidentified sources found by *Fermi*-Large Area Telescope (LAT; Ackermann et al. 2012), the claimed detection of a signal consistent with dark matter in our own galaxy, which has gained much attention recently (Daylan et al. 2014), or (once a detection has been made) in the search for dark matter in dwarf galaxies (Ackermann et al. 2011, 2014, 2015; Geringer-Sameth & Koushiappas 2011; Geringer-Sameth et al. 2015). In the recent claims, the existence of a source of  $\gamma$ -rays (over some background) is established by

\* E-mail: [s.algeri14@imperial.ac.uk](mailto:s.algeri14@imperial.ac.uk)

† Wallenberg Academy Fellow.

an LRT, but the crucial and unsolved question is not whether a  $\gamma$ -ray source exists, but whether it can be explained by conventional sources of  $\gamma$ -rays as opposed to dark matter annihilation. This is a prime example of a non-nested model comparison. For definiteness, we can assume  $f(y, E_0, \phi) \propto \phi E_0^\phi y^{-(\phi+1)}$  is the probability density function (pdf) of the  $\gamma$ -rays energies, denoted by  $y$ , originating from known cosmic sources and  $g(y, M_\chi) \propto 0.73(\frac{y}{M_\chi})^{-1.5} \exp\{-7.8 \frac{y}{M_\chi}\}$  is the pdf of the  $\gamma$ -ray energies of dark matter (Bergström, Ullio & Buckley 1998). The goal is to decide if  $f(y, E_0, \phi)$  is sufficient to explain the data ( $H_0$ ) or if  $g(y, M_\chi)$  ( $H_1$ ) provides a better fit.

Although the issue of comparing non-nested models has been addressed since the early days of modern statistics (Cox 1961, 1962, 2013; Atkinson 1970; Quandt 1974), as well as in the more recent physical literature (Pilla, Loader & Taylor 2005; Pilla & Loader 2006), a method with the desired statistical properties, easy implementation and computational efficiency in astrophysics is still lacking.

This article is arranged as follows. Section 2 reviews the LRT, Wilks's theorem and their extensions to non-regular situations. Our proposal for testing non-tested models is introduced in Section 3, validated via simulation studies in Section 4, and applied to a realistic simulation of the *Fermi*-LAT  $\gamma$ -ray satellite in Section 5. General discussion appears in Section 6.

## 2 WILKS, CHERNOFF AND TRIAL FACTORS

Let  $f(y; \alpha)$  and  $g(y, \beta)$  be pdfs of the background and signal, where  $y$  is the detected energy,  $\alpha$  and  $\beta$  are parameters. Suppose observed particles are a mixture of background and source, i.e.

$$(1 - \eta)f(y, \alpha) + \eta g(y, \beta), \quad (1)$$

where  $0 \leq \eta \leq 1$  is the proportion of signal counts. A hypothesis test can be specified as  $H_0: \eta = \eta_0$  versus  $H_1: \eta > \eta_0$ , and if  $\beta$  is known the LRT statistic by

$$T(\beta) = -2 \log \frac{L(\eta_0, \hat{\alpha}_0, -)}{L(\hat{\eta}_1, \hat{\alpha}_1, \beta)}, \quad (2)$$

where  $L(\eta, \alpha, \beta)$  is the likelihood function under (1). The numerator and denominator of (2) are the maximum likelihood achievable under  $H_0$  and  $H_1$ , respectively with  $\hat{\alpha}_0$  being the Maximum Likelihood Estimate (MLE) of  $\alpha$  under  $H_0$  and  $\hat{\alpha}_1$  and  $\hat{\eta}_1$  the MLEs under  $H_1$ . (Wilks 1938) states that when  $H_0$  is true and when testing for a one-dimensional parameter (in this case  $\eta$ ),  $T(\beta)$  is asymptotically distributed as a  $\chi_1^2$  (the subscript being the degrees of freedom). Among the regularity conditions which guarantee this result are as follows.

RC1. The models are nested, meaning that there is a full model of which both  $H_0$  and  $H_1$  are special cases.

RC2. The set of possible parameters of  $H_0$  is on the interior of that for the full model.

RC3. The full model is identifiable under  $H_0$ .

Unfortunately in practice, it is common to encounter non-regular problems. Notice for example, if  $\beta$  is known but  $\eta_0 = 0$ , RC2 does not hold. In this case, Chernoff (1954) applies; it generalizes Wilks and states that if  $H_0$  is on the boundary of the parameter space, the asymptotic distribution of  $T(\beta)$  is an equal mixture of a  $\chi_1^2$  and a Dirac delta function at 0, namely  $\frac{1}{2}\chi_1^2 + \frac{1}{2}\delta(0)$ .

Further, if  $\eta_0 = 0$  (on the boundary) and  $\beta$  is unknown, the model in (1) is not identifiable under  $H_0$  and RC3 fails. This is known in statistics as a test of hypothesis where a nuisance parameter is defined only under  $H_1$ , or ‘trial correction’ in astrophysical literature.

A solution based on theoretical result of Davies (1987) is proposed by Gross & Vitells (2010). In particular, under  $H_0$ ,  $T(\beta)$  is a random process indexed by  $\beta$ , specifically if RC2 (but not RC3) holds  $\{T(\beta), \beta \in \mathbf{B}\}$  is asymptotically a  $\chi_1^2$ -process. A natural choice of test statistic is  $\sup_\beta T(\beta)$  and Gross & Vitells (2010) provides an approximation in the limit as  $c \rightarrow \infty$  for the tail probability  $P(\sup_\beta T(\beta) > c)$ . Finally, if both RC2 and RC3 fail to hold (e.g. the important case of  $\eta_0 = 0$  with  $\beta$  unknown), we show in our Supplementary Material that because  $\{T(\beta), \beta \in \mathbf{B}\}$  is a  $\frac{1}{2}\chi_1^2 + \frac{1}{2}\delta(0)$  random process,

$$P(\sup_\beta T(\beta) > c) \approx \frac{P(\chi_1^2 > c)}{2} + E[N(c_0)|H_0]e^{-\frac{c-c_0}{2}}, \quad (3)$$

where  $E[N(c_0)|H_0]$  is the expected number of upcrossings of the  $T(\beta)$  process over the threshold  $c_0$  under  $H_0$  and  $c_0$  is chosen  $c_0 \ll c$ . (Details of how to choose  $c_0$  are given in Gross & Vitells (2010), where (3) is also asserted, but without proof.) Although this approximation holds as  $c \rightarrow \infty$ , when  $c$  is small, the right-hand side of (3) is an upper bound for  $P(\sup_\beta T(\beta) > c)$ . Thus, basing inference on (3) is valid, though perhaps conservative.

## 3 STATISTICAL COMPARISON OF NON-NESTED MODELS

Suppose we wish to compare two pdfs,  $f(y, \alpha)$  and  $g(y, \beta)$ , for which RC1 does not apply, that is the two pdfs are not special cases of a full model and do not share a parameter space. Notice that in both  $f$  and  $g$  free parameters (i.e.  $\alpha$  and  $\beta$ , respectively) are present and thus, the problem cannot be reduced to a test for simple hypotheses as in Cousins (2005), see Cox (1961) for more details. We require  $\beta$  to be one-dimensional and  $\alpha$  to lie in the interior of its parameter space. The goal is to develop a test of the hypothesis:

$$H_0 : f(y, \alpha) \text{ versus } H_1 : g(y, \beta). \quad (4)$$

Although  $f(y, \alpha)$  and  $g(y, \beta)$  are non-nested, we can construct a comprehensive model which includes both as special cases. There are two reasonable formulations. We encountered the first in (1); the second is proportional to  $\{f(y, \alpha)\}^{1-\eta}\{g(y, \beta)\}^\eta$ , with  $0 \leq \eta \leq 1$  in both formulations. As discussed in Cox (1962), Cox (2013), Atkinson (1970) and Quandt (1974), there are advantages and disadvantages to both. From our perspective, the additive form in (1) has the advantage of more appealing mathematical properties. Since no normalizing constant is involved, the maximization of the log-likelihood reduces to numerical optimization. In contrast to the test discussed in Section 2, the model in (1) is not viewed as a mixture of astrophysical models in which a certain proportion of events,  $\eta$ , originates a process represented by one model, and the remaining proportion,  $1 - \eta$ , originates from the completing process represented by the other model. Instead, (1) is a mathematical formalization used to embed the pdfs  $f(y, \alpha)$  and  $g(y, \beta)$  and their corresponding parameters spaces into an overarching model via the auxiliary parameter  $\eta$  (Quandt 1974). The overarching model has not astrophysical interpretation, but helps us reformulate the test in (4) into a suitable form, i.e.

$$H_0 : \eta = 0 \text{ versus } H_1 : \eta > 0. \quad (5)$$

Perhaps a more natural formulation of (4) would be  $H_0: \eta = 0$  versus  $H_1: \eta = 1$ . Unfortunately, neither Wilks's nor 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  $\eta$  under  $H_1$ , with  $\eta = 0$  in its interior. With indirect dark matter detection, the formulation in (5)

allows the alternative model to include both the case where dark matter and known cosmic sources are present simultaneously ( $0 < \eta < 1$ ) and the case where only dark matter is present ( $\eta = 1$ ). In situations where intermediate values of  $\eta$  are not physical we might, in addition to (5), test  $H_0: \eta = 1$  versus  $H_1: \eta < 1$ , i.e. interchange the roles of the hypotheses as discussed in Cox (1962, 2013). In this case, the nuisance parameter  $\alpha$  is required to be one-dimensional i.e.  $\alpha = \alpha$ .

Under model (1), testing (5) is equivalent to testing  $\eta$  on the boundary with  $\beta$  only being defined under the alternative. We can apply the methods discussed in Section 2 to solve this problem. Notice that such methods can still be applied if the two models share additional parameters,  $\gamma$ , i.e.  $f(y, \gamma, \alpha)$  and  $g(y, \gamma, \beta)$ . However, the maximized likelihoods in (2) must be replaced by their profile counterparts  $L(0, \hat{\gamma}_0, \hat{\alpha}_0)$  and  $L(\eta_1, \hat{\gamma}_1, \hat{\alpha}_1, \beta)$  (Davison 2003).

#### 4 VALIDATION ON DARK MATTER MODELS

We illustrate the reliability of the method proposed for testing non-nested models using two sets of Monte Carlo simulations. In Test 1, we compare the two models introduced in Section 1 with the aim of distinguishing between a dark matter signal and a power-law distributed cosmic source. In Test 2, we make the same comparison but in the presence of power-law distributed background. In this case,  $H_0$  specifies as

$$f(y, \delta, \lambda, E_0, \phi) = (1 - \lambda) \frac{\delta E_0^\delta}{k_\delta y^{\delta+1}} + \lambda \frac{\phi}{k_\phi y^{\phi+1}} E_0^\phi \quad (6)$$

and  $H_1$  specifies

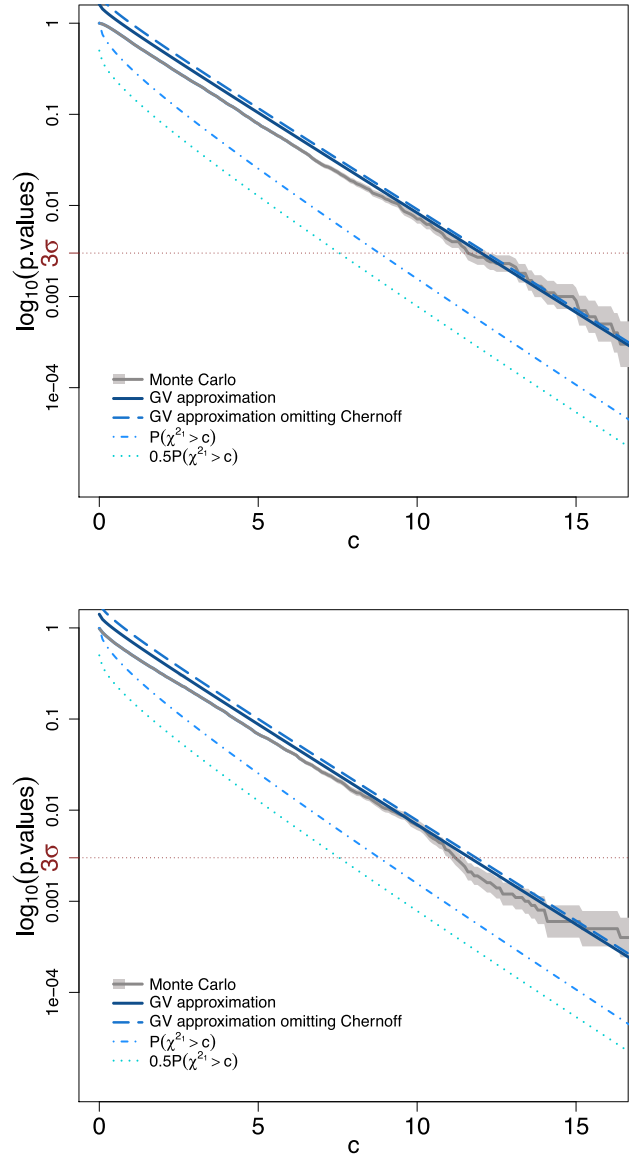
$$g(y, \delta, \lambda, E_0, M_\chi) = (1 - \lambda) \frac{\delta E_0^\delta}{k_\delta y^{\delta+1}} + \lambda \frac{e^{-7.8 \frac{y}{M_\chi}}}{y^{1.5} k_{M_\chi}}; \quad (7)$$

where  $k_\phi$ ,  $k_\delta$  and  $k_{M_\chi}$  are the normalizing constants for each pdf,  $0 < \lambda < 1$ ,  $\delta > 0$ ,  $\phi > 0$ ,  $E_0 = 1$ ,  $y \in [E_0, 100]$  and  $M_\chi \in [E_0, 100]$ . Note that in this case, the formulation in (1), with mixture parameter  $\lambda$ , is first used to specify the signal existence over a (relatively well known) background, whilst in the next step, equation (1) is adopted as a merely mathematical tool to treat the non-nested case (as described previously).

For simplicity, in Test 2,  $\lambda$ , the proportion of events coming from dark matter, was fixed to 0.2. In both tests, we estimated the average number of upcrossings  $E[N(c_0)|H_0]$  using 1000 Monte Carlo simulations. Finally, the approximation to  $P(\sup_\beta T(\beta) > c)$  is calculated using (3) on a grid of values of  $c$ . The results are compared with the respective Monte Carlo p-values. In Fig. 1 along with the  $\chi^2$  and Chernoff corrections one might compute ignoring the regularity conditions in Section 2.

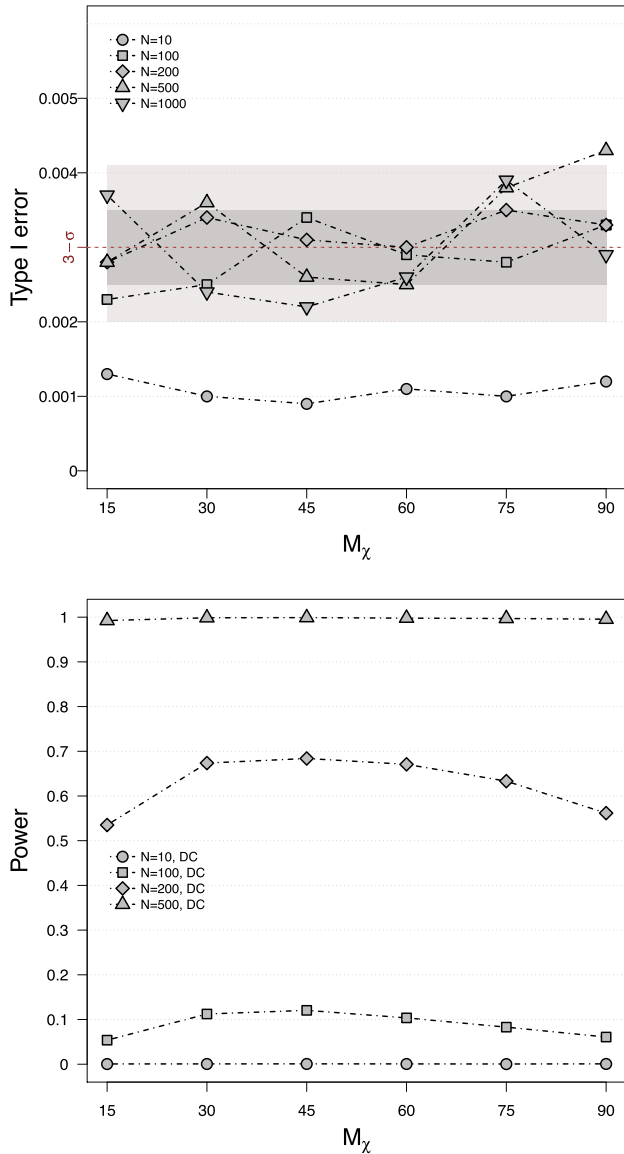
For small  $c$ , the approximation in (3) is greater than its Monte Carlo counterpart. As  $c$  increases, however, the approximation converges to the Monte Carlo estimates for a good approximation to the p-value,  $P(\sup_\beta T(\beta) > c)$ . The  $\chi^2$  and respective Chernoff-adjusted approximation lead to over optimistic p-values, whereas similar results to those attained with (3) are achieved when the factor of 2 that accounts for RC1 is omitted. This is not surprising since the right-hand side of (3), is dominated by  $E[N(c_0)|H_0]$  (which also explains the wide discrepancy between (3) and the  $\chi^2$  approximations in Fig. 1) and in practice, when testing on the boundary of the parameter space,  $E[N(c_0)|H_0]$  is typically calculated simulating a  $\frac{1}{2}\chi_1^2 + \frac{1}{2}\delta(0)$  random process directly. Thus, the Chernoff correction is automatically implemented in the leading term of (3).

It is not uncommon in practice, e.g. in astronomy, for the number of counts to be considerably smaller than the 10 000 used in Fig. 1.



**Figure 1.** Comparing the approximation in (3) (solid blue lines) with Monte Carlo estimation of  $P(\sup T(M_\chi) > c)$  (grey dashed lines), for Test 1 (upper panel) and Test 2 (lower panel). Approximations corresponding to (3) without the Chernoff correction (blue dashed lines), a  $\chi^2$  approximation (light blue dash-dotted lines) and a Chernoff-adjusted  $\chi^2$  approximation (light blue dotted lines) are also reported. Monte Carlo p-values were obtained by simulating 10 000 data sets under  $H_0$ , each of size 10 000 for both simulations. For each simulated data set  $\sup_{M_\chi} T(M_\chi)$  was computed over an  $M_\chi$  grid of size 100 for Test 1 and size 400 for Test 2. Monte Carlo errors (grey areas) were attained via error propagation (Cowan 1998).

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 (3) holds. The upper panel of Fig. 2 reports the simulated type-I errors with a detection threshold on the p-value of 0.003 ( $3\sigma$ ) for different sample sizes when conducting Test 1. For sample sizes of at least 100, the Monte Carlo results are consistent with the numerical  $3\sigma$  error rate. The lower panel of Fig. 2 shows the power (probability of detection) curves at  $3\sigma$  of the same test for different sample sizes. For all the values of  $M_\chi$  considered, a sample size of 500 is sufficient to achieve a power of nearly 1.



**Figure 2.** Simulated type-I errors (upper panel) and power functions (lower panel) for Test 1 with at  $3\sigma$  significance. Shaded areas indicate regions expected to contain 68 per cent (dark grey) and 95 per cent (light grey) of the symbols if the nominal type-I error of 0.003 holds. For both the type-I error and power curves 10 000 Monte Carlo simulations were used.

## 5 APPLICATION TO SIMULATED DATA FROM THE *FERMI*-LAT

The *Fermi*-LAT (Atwood 2009) is a pair-conversion  $\gamma$ -ray telescope on board the earth-orbiting *Fermi* satellite. It measures energies

and images  $\gamma$ -rays between about a 100 MeV and several TeV. One particular aspect is the  $\gamma$ -ray signal induced by dark matter annihilations, which gives rise to measurable signal from celestial objects, like the Milky Way centre or dwarf galaxies. Here, we apply the method proposed in this letter to a data set simulated with realistic representations of the effects of the detector and present backgrounds. We considered a 5 yr observation of putative dark matter source (dwarf galaxy-like) with dark matter annihilating into b-quark pairs and a mass of the dark matter particle of 35 GeV. This assumption is consistent with the most generic and popular models for dark matter, namely that it is in large part made of a weakly interacting massive particles. It is also consistent with recent claims of evidence for dark matter. The signal normalization corresponds to about 200 events detected in the LAT. Roughly, this corresponds to a dark matter source at the distance of the dwarf galaxy Segue 1 (and with comparable dark matter density) and an annihilation cross-section of  $\sim 2 \cdot 10^{-25} \text{ cm}^3 \text{ s}^{-1}$ ). Results are summarized in Table 1. We find a  $4.198\sigma$  significance in favour of the dark matter model. Scaling the event rate down to 50 (i.e. considering a lower cross-section by a factor of 4 or lower density by a factor of 16), we obtain  $2.984\sigma$  significance (result not shown). Adding complexity, we introduce a background, for example  $\gamma$ -rays introduced by our own Galaxy. We then considered 2176 counts from a power-law distributed background source as in (6)–(7) and about 550 dark matter events. For simplicity, the mixture parameter  $\lambda$  is fixed at 0.2. In this case, we find  $2.9\sigma$  significance in favour of the model in (7). 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 reduce background by taking  $\gamma$ -ray directions into account.

## 6 SUMMARY AND DISCUSSION

We have presented a two-step solution to a common problem in experimental astrophysics: comparing competing non-nested models. On the basis of the seminal work of Cox (1962, 2013) and Atkinson (1970), the first step of our strategy requires the specification of a comprehensive model which extends the parameter space of the models to be compared. The problem of testing non-nested model is then reduced to the look-elsewhere effect, and thus the second step naturally recalls Gross & Vitells (2010) as an efficient solution to accurately approximate the significance of new signals. The resulting procedure is easy to implement, does not require extensive calculations on a case-by-case basis and is computationally more efficient than Monte Carlo simulations. Recent developments (Algeri et al. 2016) in the nested case illustrate additional desirable statistical properties of Gross & Vitells (2010) with respect to Pilla et al. (2005) and Pilla & Loader (2006). Given the nature of the methodology proposed in this letter, we expect these finding to carry over to the non-nested case.

**Table 1.** Summary of the analysis on the *Fermi*-LAT simulation comparing the models in Tests 1 and 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  $H_1: \eta < 1$ .

	$H_0$	$N$	$\hat{\eta}$	$\hat{M}_\chi$	sup LRT	Sig.
Test 1	$\eta = 0$	200	0.971	27	21.018	$4.038\sigma$
	$\eta = 1$	200	p-value = 0.528			
Test 2	$\eta = 0$	2726	0.999	30	12.096	$2.673\sigma$
	$\eta = 1$	2726	p-value = 1			

An example of testing non-nested models arises in the search for particle dark matter. We use this example to validate and illustrate the procedure. We also demonstrate good performance in a realistic simulation of data that is collected with the *Fermi*-LAT  $\gamma$ -ray detector and used in the search for signal from dark matter annihilation.

Although any pair of hypothesized models can be expressed as a special case of (1), this formulation alone does not always provide a mechanism for a statistical hypothesis test. In the non-nested scenario analysed in this letter, valid inference for the test in (5) can be achieved by applying the methodology in Gross & Vitells (2010). This method, however, cannot handle multidimensional parameters that are defined only under  $H_1$ , nor can it deal with nuisance parameters under  $H_0$  which lie on the boundary of their parameter space. A possible approach to tackle the first limitation is to apply the theory in Vitells & Gross (2011) to the comprehensive model in (1). Whereas an extension of the method to overcome the second limitation could rely on the theory in Self & Liang (1987). In light of this, the methodology proposed is particularly suited to comparisons of non-nested models where these limitations often do not arise.

Software for the methodology illustrated in this letter is available at <http://www.imperial.ac.uk/sa2514/Research.html>.

## ACKNOWLEDGEMENTS

The authors acknowledge Brandon Anderson for using tools publicly available from the *Fermi*-LAT Collaboration to simulate *Fermi*-LAT data. JC thanks the support of the Knut and Alice Wallenberg foundation and the Swedish Research Council. DvD acknowledges support from a Wolfson Research Merit Award provided by the British Royal Society and from a Marie-Curie Career Integration Grant provided by the European Commission.

## REFERENCES

- Ackermann M. et al., 2011, *Phys. Rev. Lett.*, 107, 241302  
 Ackermann M. et al., 2012, *ApJ*, 747, 121  
 Ackermann M. et al., 2014, *Phys. Rev. D*, 89, 042001  
 Ackermann M. et al., 2015, *Phys. Rev. Lett.*, 115, 231301  
 Algeri S., van Dyk D. A., Conrad J., Anderson B., 2016, preprint ([arXiv:1602.03765v2](https://arxiv.org/abs/1602.03765v2))  
 Atkinson. A. C. 1970, *J. R. Stat. Soc. B*, 32, 3  
 Atwood W. B. et al., 2009, *ApJ*, 697, 2  
 Bergström L., Ullio P., Buckley J., 1998, *Astropart. Phys.*, 9, 2  
 Chernoff H., 1954, *Ann. Math. Stat.*, 25, 3  
 Conrad J., 2015, *Astropart. Phys.*, 62, 165  
 Conrad J., Cohen-Tanugi J., Stigari L. E., 2015, *J. Exp. Theor. Phys.*, 148, 12  
 Cousins R., Mumford J., Tucker J., Valuev V., 2005, *J. High Energy Phys.*, 11, 046  
 Cowan G., 1998, *Statistical Data Analysis*. Oxford Univ. Press, Oxford  
 Cox D. R., 1961, in Neyman J., ed., *Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, Contributions to the Theory of Statistics. Univ. California Press, Berkeley, p. 105  
 Cox D. R., 1962, *J. R. Stat. Soc. B*, 24, 2  
 Cox D. R., 2013, *J. R. Stat. Soc. B*, 75, 2  
 Davies R. B., 1987, *Biometrika*, 74, 1  
 Davison A. C., 2003, *Statistical Models*. Cambridge Univ. Press, Cambridge  
 Daylan T., Finkbeiner D. P., Hooper D., Linden T., Portillo S. K. N., Rodd N. L., Slatyer T. R., 2014, preprint ([arXiv:1402.6703](https://arxiv.org/abs/1402.6703))  
 Geringer-Sameth A., Koushiappas S. M., 2011, *Phys. Rev. Lett.* 107, 241303  
 Geringer-Sameth A., Walker M. G., Koushiappas S. M., Koposov S. E., Belokurov V., Torrealba G., Evans N. W., 2015, *Phys. Rev. Lett.*, 115, 081101  
 Gross E., Vitells O., 2010, *Eur. Phys. J. C*, 70, 1  
 Guiriec S. et al., 2015, *ApJ*, 807, 2  
 Pilla R., Loader C., 2006, preprint ([arXiv:math/0511503v2](https://arxiv.org/abs/math/0511503v2))  
 Pilla R., Loader C., Taylor C. C., 2005, *Phys. Rev. Lett.*, 95, 230202  
 Profumo S., Linden T., 2012, *J. Cosmol. Astropart. Phys.*, 07, 011  
 Quandt R. E., 1974, *Rev. Econ. Stat.*, 5, 1  
 Self S. G., Liang K.-Y., 1987, *J. Am. Stat. Assoc.*, 82, 398  
 Vitells O., Gross E., 2011, *Astropart. Phys.*, 35, 5  
 Wilks S. S., 1938, *Ann. Math. Stat.*, 9

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

### suppl\_data

(<http://www.mnrasl.oxfordjournals.org/lookup/suppl/doi:10.1093/mnrasl/slw025/-/DC1>).

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.