University of Amsterdam

UvA-DARE (Digital Academic Repository)

A fresh approach to forecasting in astroparticle physics and dark matter searches

Edwards, T.D.P.; Weniger, C.

Link to publication

A Fresh Approach to Forecasting in Astroparticle Physics and Dark Matter Searches

Thomas D. P. Edwards and Christoph Weniger

Gravitation Astroparticle Physics Amsterdam (GRAPPA), Institute for Theoretical Physics Amsterdam and Delta Institute for Theoretical Physics, University of Amsterdam, Science Park 904, 1090 GL Amsterdam, The Netherlands

E-mail: t.d.p.edwards@uva.nl, c.weniger@uva.nl

Abstract. We present a toolbox of new techniques and concepts for the efficient forecasting of experimental sensitivities. These are applicable to a large range of scenarios in (astro-)particle physics, and based on the Fisher information formalism. Fisher information provides an answer to the question 'what is the *maximum* extractable information from a given observation?'. It is a common tool for the forecasting of experimental sensitivities in many branches of science, but rarely used in astroparticle physics or searches for particle dark matter. After briefly reviewing the Fisher information matrix of general Poisson likelihoods, we propose very compact expressions for estimating expected exclusion and discovery limits ('equivalent counts method'). We demonstrate by comparison with Monte Carlo results that they remain surprisingly accurate even deep in the Poisson regime. We show how correlated background systematics can be efficiently accounted for by a treatment based on Gaussian random fields. Finally, we introduce the novel concept of Fisher *information flux*. It can be thought of as a generalization of the commonly used signal-to-noise ratio, while accounting for the non-local properties and saturation effects of background and instrumental uncertainties. It is a powerful and flexible tool ready to be used as core concept for informed strategy development in astroparticle physics and searches for particle dark matter.

Contents

1 Introduction

Progress in astroparticle physics and dark matter (DM) searches is driven by the comparison of theoretical models with experimental data. During this process, estimating the sensitivity of existing or future experiments for the detection of astrophysical or new physics signals is a ubiquitous task, usually requiring the calculation of the expected exclusion and discovery limits. It is done both by phenomenologists who are interested in observational prospects of their theoretical models, as well as experimentalists who are interested in optimizing

experimental design. As such, efficient and informative forecasting plays a central role in shaping the development of the field. It is rather common in the (astro-)particle physics community that forecasting is done in the framework of Frequentist statistics. Acceptance of the signal+background hypothesis, H_1, and rejection of the background-only hypothesis, H_0, requires that some test statistic (TS) exceeds a predefined threshold, which depends both on the aspired significance level of the detection as well as the probability distribution function (PDF) of the TS under H_0 [1, 2]. The Frequentist method (in contrast to Bayesian techniques) has the advantage of known false positive rates for hypothesis testing, and known coverage for upper limits. These features can be especially desirable when prior knowledge about the (non-)realization of a theoretical models in Nature is poor.

One of the most commonly used TSs (used for parameter regression, calculation of confidence intervals, and the goodness-of-fit) is the *chi-squared*. Its application is limited to binned data with errors that are approximately normal distributed, which is often realized in the large-sample limit. Asymptotic formulae for its statistical properties are available [3]. In the small sample limit, and generally for more complicated likelihood functions, the maximum likelihood ratio (MLR) [3] is a very common TS. Again, in the large sample limit, asymptotic distributions are available [4, 5]. In the small-sample limit, however, its statistical properties have to be inferred from Monte Carlo (MC) simulations.

When estimating experimental sensitivities, one general question that arises is '*What is the maximum information that can be in principle extracted from a given observation?*'. Information gain here corresponds to the reduction of the uncertainty associated to the model parameters of interest. The famous Cramér-Rao bound [e.g. 2] provides a general way to derive, for a given experimental setup, a lower bound on the achievable variance of any unbiased model parameter estimator, which holds for *any* statistical method employed to analyse the data. This bound corresponds to an upper limit on the achievable information gain. The Cramér-Rao bound is based on the *Fisher information matrix*, which quantifies how 'sharply peaked' the likelihood function describing the observational data is around its maximum value.

Forecasting instrumental sensitivities with Fisher information is rather common, *e.g.*, in cosmology. It rests on the assumption that estimators which saturate the Cramér-Rao bound are available, and that these estimators approximately follow a multi-variate normal distribution. In the large-sample limit, this is indeed often (but not always) the case. The important limitations of this approach were pointed out many times [*e.g.*, 6, 7], as well as proposals to extend the framework to account for, *e.g.*, non-Gaussian effects [8] (for a recent review see Ref. [9]).

The Fisher information matrix has an impressive range of attractive properties that – if used wisely – can significantly ease the life of anybody interested in performing forecasting, from simple tasks to problems with many experiments, targets and a high-dimensional parameter space. It can be quickly calculated, it is additive, allows for efficient handling of nuisance parameters, and it is at the root of the powerful concept of information geometry [*e.g.*, 10] (and, as we will show in this work, *information flux*). Fisher forecasting is largely unused in the astroparticle physics and DM communities however there are some exceptions, *e.g.* [6, 10–15].

In this paper, we present an overview of how Fisher information can be used in the context of astroparticle physics and DM searches. Throughout this paper, we focus our attention on Poisson likelihoods and *additive component models*, where the shape (in for

instance energy or positional space) of the model components are fixed and the normalization coefficients are the only regression parameters. In addition, we will study the impact of additional *external constraints* on the model parameters. Such constraints can be used to account for various model or instrumental uncertainties. Focusing on these scenarios allows us an in-depth discussion of the specific capabilities and limitations of Fisher forecasting, while still covering many interesting use cases.

We introduce various (to the best of our knowledge) new prescriptions for the efficient estimate of the expected exclusion and discovery limits that are valid in the small- and large-sample regimes based on the Fisher information matrix. We compare the accuracy of these prescriptions with results from the MLR method, and for a few simple cases with results from the full Neyman belt construction. We demonstrate how to incorporate the effect of correlated background systematics. Lastly, we introduce the new concept of Fisher *information flux*. It generalizes the commonly used signal-to-noise ratio (SNR), while fully accounting for background and other uncertainties. We illustrate the power of this new concept in a few simple examples.

Some of the techniques discussed in this paper are extensively used in other fields of science. Part of the work is inspired by the discussion of information geometry in Ref. [10], and by the notion of effective backgrounds in Refs. [14, 16]. We mention explicitly when and why we deviate from notations introduced in these works if we deem this to be necessary. Some of the methods discussed here have already been applied by some of the present authors in Refs. [12, 13].

The paper is organized as follows: In Sec. 2, we introduce and define the Fisher information matrix, additive component models, and equivalent signal and background counts. In Sec. 3, we present prescriptions to derive approximate expected exclusion limits and discovery sensitivities from the Fisher information matrix, and study the validity of the results with MC techniques. In Sec. 4 we show in a few examples how systematic background uncertainties can be incorporated in the sensitivity estimates. In Sec. 5 we introduce the notion of Fisher information flux, and various connected concepts. In Sec. 6 we finally conclude.

In Appendix A we discuss relevant properties of the Poisson likelihood function and the associated Fisher information. In Appendix B, we discuss conventional methods for the calculation of the discovery reach and expected upper limits, that we use for comparison with our techniques. Finally, in Appendix C we collect some more technical derivations for results in the paper.

2 Fisher Information of the Poisson likelihood function

2.1 The Fisher information matrix

For any sufficiently regular likelihood function $\mathcal{L}(\mathcal{D}|\boldsymbol{\theta})$, with n model parameters $\boldsymbol{\theta} \in \mathbb{R}^n$ and data $\mathcal{D}$, the Fisher information matrix is a $n \times n$ matrix that can be defined as

$$\mathcal{I}_{ij}(\boldsymbol{\theta}) \equiv \left\langle \left(\frac{\partial \ln \mathcal{L}(\mathcal{D}|\boldsymbol{\theta})}{\partial \theta_i} \right) \left(\frac{\partial \ln \mathcal{L}(\mathcal{D}|\boldsymbol{\theta})}{\partial \theta_j} \right) \right\rangle_{\mathcal{D}(\boldsymbol{\theta})} = - \left\langle \frac{\partial^2 \ln \mathcal{L}(\mathcal{D}|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right\rangle_{\mathcal{D}(\boldsymbol{\theta})} , \qquad (2.1)$$

where the average is taken over multiple realizations of a model with parameters $\boldsymbol{\theta}$. The second equality holds given some weak regularity conditions [17]. The Fisher information matrix quantifies the *maximum precision* at which the model parameters can be inferred from

the data.[1] This follows from the Cramér-Rao bound (CRB) [18, 19], which states that for *any* set of unbiased estimators of the model parameters, $\hat{\boldsymbol{\theta}}$, the inverse of the Fisher information matrix provides a lower limit on its variance. The CRB generalises to the multivariate case to give [1],

$$\mathrm{cov}\left[\hat{\theta}_i, \hat{\theta}_j\right] \equiv \langle(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)\rangle_{\mathcal{D}(\boldsymbol{\theta})} \geq (\mathcal{I}(\hat{\boldsymbol{\theta}})^{-1})_{ij} \,. \tag{2.2}$$

Estimators that saturate the bound exactly are called 'minimum variance'. An estimator is called 'asymptotically efficient' when the bound is saturated in the large sample limit. Note that this bound only holds for *unbiased* estimators, and becomes stronger in the presence of a constant bias [1].

A widely used estimator is the maximum likelihood estimator (MLE). The MLE is *exact*, meaning that although it is in general biased it becomes unbiased in the large sample limit. Furthermore, the MLE is asymptotically efficient. Although the Fisher information matrix as an estimate for the variance of MLEs becomes exact only in the large-sample limit, we will see below that it remains a powerful tool in the small-sample regime.

Searches for new physics in both astrophysics and DM detection experiments are often, at their core, counting experiments. The number of events recorded in a detector is described by the Poisson distribution. We assume here for simplicity that events can be described by one variable, for example photon energy E (the generalization to multiple variables is straightforward and will be used below). Then, the Poisson log-likelihood can be written as (details can be found in Appendix A.1)

$$\ln \mathcal{L}_{\mathrm{pois}}(\mathcal{D}|\boldsymbol{\theta}) = \int dE \, \left[\mathcal{C}(E) \ln \Phi(E|\boldsymbol{\theta}) - \Phi(E|\boldsymbol{\theta})\right] \,, \tag{2.3}$$

where the integration is done over the energy range of interest, and we dropped terms that do not depend on $\boldsymbol{\theta}$ since they do not affect the rest of the discussion. Here, $\Phi(E|\boldsymbol{\theta})$ is the model counts spectrum (with units $1/E$). Furthermore, we defined the 'unbinned' count spectrum

$$\mathcal{C}(E) = \sum_k \delta_D(E - E_k) \,, \tag{2.4}$$

where E_k is the energy of event k, and $\delta_D(E)$ is the Dirac delta function. $\mathcal{C}(E)$ can be interpreted as a counts histogram with zero bin size. It represents a specific realization of a Poisson process with mean $\Phi(E|\boldsymbol{\theta})$, and has the useful property that $\langle\mathcal{C}(E)\rangle_{\mathcal{D}(\boldsymbol{\theta})} = \Phi(E|\boldsymbol{\theta})$.

We note that the above definition resembles a *Poisson point process*, defined here on an interval on the real line that is given by the energy range of interest [20, 21]. In this context, $\Phi(E|\boldsymbol{\theta})$ would be referred to as 'intensity measure', but we will continue to use here the expressions 'model counts spectrum' or, in some 2-dim examples below, 'model counts map'.

It is straightforward to show that the Fisher information matrix of the above Poisson likelihood is given by

$$\mathcal{I}_{ij}^{\mathrm{pois}}(\boldsymbol{\theta}) = \int dE \, \frac{\partial \Phi(E|\boldsymbol{\theta})}{\partial \theta_i} \frac{1}{\Phi(E|\boldsymbol{\theta})} \frac{\partial \Phi(E|\boldsymbol{\theta})}{\partial \theta_j} \,. \tag{2.5}$$

---

[1]Here, we refer to the maximum precision in terms of a Frequentist approach. In the Bayesian case it is possible to gain further insight with the addition of informative priors and/or a hierarchical model structure.

This form is fully general and holds for any parametric dependence of $\Phi(E|\boldsymbol{\theta})$. The inverse of the Fisher information matrix (irrespectively of whether it is derived from Poisson or other likelihood functions) approximates the expected covariance matrix as shown in Eq. (2.2). As mentioned above, this relation holds exactly in the large-sample limit only.

For the diagonal elements of the inverse of the Fisher matrix, we will sometimes use of the notation

$$\sigma_i^2(\boldsymbol{\theta}) \equiv (\mathcal{I}(\boldsymbol{\theta})^{-1})_{ii} \,. \tag{2.6}$$

Furthermore, we note that in general the full Fisher matrix can have, beside the Poisson part that we discussed above, parts that introduce additional constraints on the model parameters (see Sec. 4 below). In that case, we can write

$$\mathcal{I}_{ij}(\boldsymbol{\theta}) = \mathcal{I}_{ij}^{\text{pois}}(\boldsymbol{\theta}) + \mathcal{I}_{ij}^{\text{syst}} \,, \tag{2.7}$$

where we assumed that the systematics term is (approximately) independent of the model parameters.

2.2 The profiled Fisher information matrix

Typically, one is only interested in a few 'parameters of interest' (PoI), say $\theta_1, \ldots, \theta_k$, while the remaining $n - k$ model parameters are nuisance parameters that parametrize background, instrumental or signal uncertainties. The canonical method to deal with MLEs of nuisance parameters in a Frequentist approach is to maximise the likelihood function $\mathcal{L}(\mathcal{D}|\boldsymbol{\theta})$ with respect to the parameters $\theta_{k+1}, \ldots, \theta_n$, which gives rise to a 'profile likelihood' function that only depends on the k PoI. This method leaves one with a description of the remaining parameter space whilst accounting for the effects of the $n - k$ nuisance parameters.

To perform the analogous procedure for the Fisher information matrix, we write $\mathcal{I}_{ij}(\boldsymbol{\theta})$ in block form,

$$\mathcal{I} = \begin{pmatrix} \mathcal{I}_A & \mathcal{I}_C^T \\ \mathcal{I}_C & \mathcal{I}_B \end{pmatrix} \,, \tag{2.8}$$

where $\mathcal{I}_A$ is a $k \times k$ matrix that corresponds to the PoI, $\mathcal{I}_B$ is a $(n-k) \times (n-k)$ matrix that corresponds to the nuisance parameters, and $\mathcal{I}_C$ is the mixing between both. We then define the *profiled Fisher information matrix*, where the nuisance parameters are removed in such a way that the (co-)variance of the PoIs remains unchanged. It is given by[2]

$$\widetilde{\mathcal{I}}_A = \mathcal{I}_A - \mathcal{I}_C^T \mathcal{I}_B^{-1} \mathcal{I}_C \,. \tag{2.9}$$

This expression is general and does not depend on the details of the problem.

2.3 Additive component models

To simplify the discussion we will, as mentioned above, assume that the model counts spectrum $\Phi(E|\boldsymbol{\theta})$ consists of a number of additive components with free normalization θ_i but fixed shape $\Psi_i(E)$,

$$\Phi(E|\boldsymbol{\theta}) = \sum_{i=1}^{n} \theta_i \Psi_i(E) \,, \tag{2.10}$$

---

[2]This result can be readily understood by acknowledging that the upper left $k \times k$ part of the inverse of Eq. (2.8) is given by $(\mathcal{I}_A - \mathcal{I}_C^T \mathcal{I}_B^{-1} \mathcal{I}_C)^{-1}$.

where the sum is taken over the n model components. The model counts spectra can be in many cases calculated as

$$\Psi_i(E) = \mathcal{E}(E)\, I_i(E) \,, \tag{2.11}$$

where $\mathcal{E}(E)$ denotes the exposure (usually effective area or volume times effective observation time) as function of energy, and $I_i(E)$ the differential flux of component i. We furthermore will occasionally use the total flux $I(\boldsymbol{\theta}|\Omega) \equiv \sum_i \theta_i I_i(\Omega)$. The number of expected counts per component is given by

$$\lambda_i(\theta_i) \equiv \theta_i \int dE\, \Psi_i(E) \,. \tag{2.12}$$

For the additive component model, the Fisher information matrix acquires the simple form

$$\mathcal{I}_{ij}^{\mathrm{pois}}(\boldsymbol{\theta}) = \int dE \frac{\Psi_i(E)\Psi_j(E)}{\Phi(E|\boldsymbol{\theta})} \,. \tag{2.13}$$

Note that the model parameters enter here only through the model predictions in the denominator.

2.4 Equivalent number of signal and background events

In Sec. 3 we will present methods for calculating approximate upper limits and discovery reaches which rely on the definitions of the number of signal and background counts. We refer to the method as the Equivalent Counts Method (ECM) since it is a generalisation of the single binned case where it is clear that the number of signal and background events provides information about (a) the expected significance of the signal, (b) the signal-to-background ratio (SBR) and (c) the sample size and hence the relevance of the discreteness or skewness of the Poisson distribution. The ECM definitions capture the same information but for more general cases i.e. situations with large numbers of bins or even the unbinned case as is presented here. In general, not all signal events $\lambda_i(\theta_i)$ are statistically relevant, since some might overlap with regions containing strong backgrounds, whilst other signal events might reside in regions that are almost background free. A reasonable definition of equivalent signal and background events should account for this effect.

We propose definitions here for the number of equivalent signal and background counts that are defined *solely* in terms of the Fisher information matrix. They provide information about the expected signal significance, the SBR and the effective sample size in rather general situations. These definitions are used below for the calculation of expected exclusion and discovery limits.

We define the equivalent signal, $s_i(\boldsymbol{\theta})$, and background, $b_i(\boldsymbol{\theta})$, events for any component i implicitly in terms of the SNR,

$$\mathrm{SNR}_i(\boldsymbol{\theta}) = \frac{s_i^2(\boldsymbol{\theta})}{s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta})} \,, \tag{2.14}$$

and in terms of the SBR,

$$\mathrm{SBR}_i(\boldsymbol{\theta}) = \frac{s_i(\boldsymbol{\theta})}{b_i(\boldsymbol{\theta})} \,. \tag{2.15}$$

The SNR for model component i is given by the corresponding diagonal term of the inverse of the Fisher information matrix, times the factor θ_i^2, which we can write as

$$\mathrm{SNR}_i(\boldsymbol{\theta}) \equiv \frac{\theta_i^2}{\sigma_i^2(\boldsymbol{\theta})} \,. \tag{2.16}$$

If systematic (non-Poisson) contributions to the $\mathcal{I}_{ii}(\boldsymbol{\theta})$ as well as mixing with other components are negligible, $\mathrm{SNR}_i(\boldsymbol{\theta})$ is essentially given by the corresponding diagonal component of Eq. (2.13). As expected, this is simply the standard SNR of $\theta_i\Psi_i(E)$ over a fixed background $\Phi(E|\boldsymbol{\theta})$. The definition of the SBR in terms of the Fisher information matrix is less obvious. We find that the following expression serves the purpose,

$$\mathrm{SBR}_i(\boldsymbol{\theta}) \equiv \frac{\sigma_i^2(\boldsymbol{\theta})}{\sigma_i^2(\boldsymbol{\theta}_0)} - 1 \ . \tag{2.17}$$

Here, we used the definition $\boldsymbol{\theta}_0 \equiv (\theta_1, \ldots, \theta_{i-1}, 0, \theta_{i+1}, \ldots, \theta_n)^T$, *i.e.* it is $\boldsymbol{\theta}$ with the normalization of the signal component set to zero. This choice will be further justified below.

From the definitions in Eqs. (2.16) and (2.17), the equivalent number of signal and background events can be directly obtained using Eqs. (2.14) and (2.15). They are given by

$$s_i(\boldsymbol{\theta}) = \frac{\theta_i^2}{\sigma_i^2(\boldsymbol{\theta}) - \sigma_i^2(\boldsymbol{\theta}_0)} \ , \tag{2.18}$$

and

$$b_i(\boldsymbol{\theta}) = \frac{\theta_i^2\sigma_i^2(\boldsymbol{\theta}_0)}{(\sigma_i^2(\boldsymbol{\theta}) - \sigma_i^2(\boldsymbol{\theta}_0))^2} \ . \tag{2.19}$$

Discussion. Although these definitions are relatively abstract, they have a number of useful properties. These become particularly clear when mixing between the components is negligible, *i.e.* $\sigma_i^2(\boldsymbol{\theta}) \simeq 1/\mathcal{I}_{ii}^{\mathrm{pois}}(\boldsymbol{\theta})$. If component shapes are very similar, such a situation can be enforced by introducing constraint terms for all non-signal components, in the way that we will discuss below in Sec. 4.1. If *component mixing is negligible*, we find:

(i) If all $\Psi_i(E)$ are constant, which corresponds to a single-bin scenario, the number of equivalent signal events just equals the total number of signal events, $s_i(\boldsymbol{\theta}) = \lambda_i(\boldsymbol{\theta})$, and the equivalent background events equals the sum of the events from all other components, $b_i(\boldsymbol{\theta}) = \sum_{j \neq i} \lambda_j(\boldsymbol{\theta})$.

(ii) In the large-signal limit, here defined as $\theta_i\Psi_i(E) \gg \sum_{j \neq i} \theta_j\Psi_j(E)$, the number of equivalent signal events of component i equals the total number of events in the signal component, $s_i(\boldsymbol{\theta}) \simeq \lambda_i(\theta_i)$. In other cases, the equivalent number of signal events is in general smaller, $s_i(\boldsymbol{\theta}) \leq \lambda_i(\theta_i)$.

(iii) The definition of the SBR for model component i in Eq. (2.17) becomes clearer when writing it in terms of the additive components functions $\Psi_i(E)$ and their sum $\Phi(E)$. In the small-signal regime, $\theta_i\Psi_i(E) \ll \sum_{j \neq i} \theta_j\Psi_j(E)$, this yields

$$\mathrm{SBR}_i(\boldsymbol{\theta}) \simeq \left(\int dE \, \frac{\Psi_i(E)^2}{\Phi(E|\boldsymbol{\theta}_0)} \right)^{-1} \int dE \, \frac{\Psi_i(E)^2}{\Phi(E|\boldsymbol{\theta}_0)} \, \frac{\theta_i\Psi_i(E)}{\Phi(E|\boldsymbol{\theta}_0)} \ . \tag{2.20}$$

Here, the ratio $\theta_i\Psi_i(E)/\Phi(E)$ can be interpreted as the SBR of component i as function of E, and $\Psi_i^2(E)/\Phi(E)$ is proportional to the SNR as function of E. Thus, Eq. (2.20) effectively averages the SBR over regions of E where the signal component $\Psi_i(E)$ is most significant w.r.t. the background $\Phi(E)$. In other words, regions in E where the background $\Phi(E)$ is intense enough that the signal $\Psi_i(E)$ is swamped do not contribute to Eq. (2.20), even if the number of signal photons in that region is high.
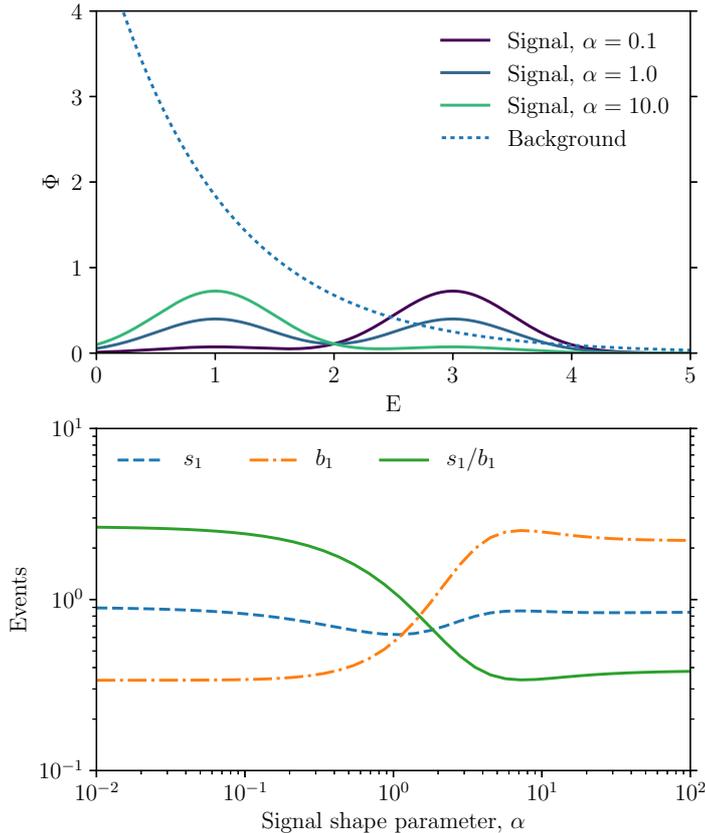
Figure 1. *Upper panel:* Example spectra of Eq. (2.22) for various values of α. *Lower panel:* Illustration of equivalent signal counts s_1, background counts b_1, and SBR s_1/b_1 as function of the signal shape parameter α.

(iv) On the other hand, in the large-signal limit (as defined above), the SBR of component i becomes

$$\mathrm{SBR}_i(\boldsymbol{\theta}) \simeq \left(\int dE \, \Psi_i(E) \right)^{-1} \int dE \, \Psi_i(E) \, \frac{\theta_i \Psi_i(E)}{\Phi(E|\boldsymbol{\theta}_0)} \, . \tag{2.21}$$

The weighting by the significance of the signal is now replaced by a weighting over the signal strength.[3]

In the *general case with mixing* between the components, the above simple analytic expressions do not hold anymore. If we concentrate on Fisher information matrices of the from Eq. (2.7), where the non-Poisson part is independent of the model parameters, one can however still show that $\mathrm{SBR}_i(\boldsymbol{\theta}) \geq 0$, which implies the important property that $b_i(\boldsymbol{\theta}) \geq 0$ and $s_i(\boldsymbol{\theta}) \geq 0$ in all cases. We confirmed this numerically by calculating SBR_i for a large number of randomly generated models.

---

[3]Eq. (2.21) is underlying the definition of the 'effective background', $b_{\mathrm{eff}} = n_s^2/\mathrm{TS}$, used in Refs. [14, 16], if we identify $n_s = \lambda_i$ and $\mathrm{TS} = \mathrm{SNR}_i$ (see Eq. (2.16)). We find that this definition can significantly underestimate the SBR in cases where most of the signal events are located in regions where they are statistically not significant. This can happen for instance in low-energy tails of a steeply falling astrophysical spectra. Our definitions based on Eq. (2.17) do not exhibit this problem.

Illustration. Finally, we illustrate the definitions introduced in this subsection with a simple example. Consider the following model

$$\Phi(E) = \theta_1 \left(\frac{\alpha}{1+\alpha} N_1(E) + \frac{1}{1+\alpha} N_3(E) \right) + \theta_2 e^{-E} \,, \tag{2.22}$$

where α is a shape parameter of the signal component, $N_1(E)$ and $N_3(E)$ are normal distributions with variance 0.5 that are centered on the indexed values, and we set the signal normalization to $\theta_1 = 1$ and the background normalization to $\theta_2 = 8$. Example spectra are shown in the upper panel of Fig. 1, for various values of α. We neglect mixing between the components by assuming that the background is fixed via external constraints.

In the lower panel of Fig. 1 we show the corresponding equivalent signal and background counts of component $i = 1$, as well as the corresponding SBR s_1/b_1. For $\alpha \ll 1$, the signal is dominated by the high energy peak, which is in a region of low background. Indeed, we find approximately $b_1 \simeq 0.33$ and $s_1 \simeq 0.89$. On the other hand, for $\alpha \gg 1$, the signal is present at lower energies, where the background is larger. Indeed, we find here $b_1 \simeq 2.2$.

In the transition region, the equivalent number of signal events s_1, which otherwise remains close to one, drops somewhat. This is expected, since at $\alpha \sim 1$ half of the signal is in a region of large backgrounds, whereas the other half is in a region with low backgrounds. Since the low-background component dominates the signal-to-noise ratio in that case, it is also mostly this component that contributes to the equivalent signal number counts. The effect becomes more pronounced if the background below $E < 2$ is further increased, or above $E > 2$ further reduced.

3 Expected exclusion limits and discovery reach

In this section, we introduce prescriptions for deriving expected exclusion and discovery limits from the Fisher information matrix. These prescriptions build on the equivalent signal and background counts that we defined above. We refer to the corresponding prescriptions as Equivalent Counts Method (ECM). For simplicity, we take the model component of interest to be $i = 1$. We put particular emphasis on the case of very small or a vanishing number of background events, and validate the approach for a few examples using MC simulations. Caveats are discussed in the last subsection.

3.1 Expected exclusion limits

Equivalent counts method. Projected approximate upper limits on the parameter θ_1, assuming that the true value is $\theta_1 = 0$, can be derived from the Fisher information matrix by solving the equation

$$s_1(\boldsymbol{\theta}^U) = Z(\alpha) \cdot \sqrt{s_1(\boldsymbol{\theta}^U) + b_1(\boldsymbol{\theta}^U)} \,, \tag{3.1}$$

for θ_1^U, while keeping the remaining $n-1$ parameters fixed to their respective values. Here, we defined $\boldsymbol{\theta}^U = (\theta_1^U, \theta_2, \ldots, \theta_n)^T$. Furthermore, s_i and b_i refer to the equivalent signal and background counts that we introduced above in Sec. 2.4. Finally, $Z(\alpha)$ is connected to the desired confidence level of the limit, $100(1-\alpha)\%$ CL, and α is the significance level. It is derived from the inverse of the standard normal cumulative distribution distribution, denoted $F_\mathcal{N}$, as

$$Z(\alpha) = F_\mathcal{N}^{-1}(1-\alpha) \,. \tag{3.2}$$
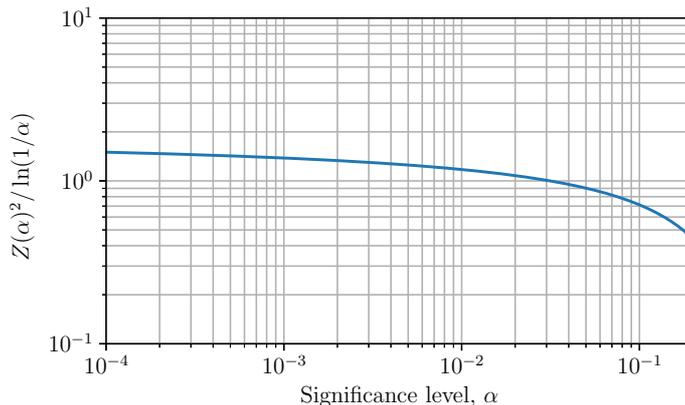
Figure 2. Ratio between expected exclusion limits from the ECM and the exact Neyman belt construction, assuming a single bin with vanishing expected background, as function of the significance level of the limit α. See Eq. (3.5) and text for details.

In the case of, say, a 95% CL upper limits, we have $\alpha = 0.05$ and hence $Z = 1.64$ [3]. It is convenient to rewrite Eq. (3.1) as

$$\theta_1^U = Z(\alpha) \cdot \sigma_1\left(\boldsymbol{\theta}^U\right) , \qquad (3.3)$$

which is in practice much easier to evaluate than Eq. (3.1).

In the *background-limited regime*, $\theta_1^U \Psi_1(E) \ll \Phi(E|\boldsymbol{\theta})$, Eq. (3.3) implies that the exclusion limit satisfies

$$\theta_1^U \simeq Z(\alpha) \cdot \sigma_1(\boldsymbol{\theta}) , \qquad (3.4)$$

where $\boldsymbol{\theta}_0 = (0, \theta_2, \ldots, \theta_n)^T$. This is exactly what is expected if Gaussian background noise with variance σ_1^2 dominates. In the *signal-limited regime*, here defined as $\theta_1^U \Psi_1(E) \gg \sum_{i \geq 2} \theta_i \Psi_i(E)$, Eq. (3.3) implies on the other hand

$$\lambda_1(\theta_1^U) \simeq Z(\alpha)^2 . \qquad (3.5)$$

The upper limit is here independent of the signal model spectrum, $\Psi_1(E)$. Eq. (3.5) should be compared with $\lambda_1 = \ln 1/\alpha$, which is the proper upper limit on the mean of a Poisson process when zero events are observed and the expected background is negligibly small. We show in Fig. 2 that the fractional difference between $Z^2(\alpha)$ and $\ln(1/\alpha)$ is indeed small for typical significance level values used to set upper limits in the literature. Namely, for $\alpha = 10^{-3}$–10^{-1} the deviation is smaller than 38%.

We note that the general method to estimate expected upper limits in Eq. (3.1) works very well even in the presence of parameter mixing and background systematics as discussed in Sec. 4, as long as the associated changes of the background flux remain 'sufficiently small' (say, below a few tens of percent) in the signal region. A quantitative discussion can be found in subsection 3.3 below.

Comparison with exact methods. The most general exact method for deriving upper limits with the correct coverage, and actually any sort of confidence intervals, is based on the Neyman Belt construction [22] (an instructive overview can be found in Ref. [23]). In

practice, often the more specific MLR method is used to construct confidence intervals of any sort. In the small-sample limit, MC simulations are required to establish the statistics of the MLR and construct intervals with the desired coverage (this is what we do here in all cases). Details about the construction of confidence intervals using both methods can be found in Appendix B.

As specific examples, we consider the signal and background functions shown in Fig. 3. Model 1 is a simple single-bin example, whereas model 2 is a basic two-bin example with a strong difference in the expected background counts in both bins. In both cases, we assume that the background normalization is known and fixed. Model 3 provides an example in the small-sample regime, and model 4 in the large-sample regime. The background normalization is here assumed to be determined by a fit to the data and hence free. Model 5 is a scenario with multiple background components that are to some degree (but not completely) degenerate with the signal.

In Fig. 4, we show the expected 95% CL exclusion limits that we obtain from our EC method, Eq. (3.3), for the five example models. We compare these limits with the median limits obtained from the MLR method. In three of the five cases (model 1, 4 & 5) the agreement is remarkably good. To emphasize this, we also indicate the extent of the 68% containment band of the MLR limits, which contains 68% of the upper limits when multiple realizations of the data are considered. They are significantly wider than the difference between the limits from the EC and the MLR methods.

It is instructive to discuss the single-bin model 1 in more detail. In Fig. 5, we show the projected 95% CL upper limits as derived from (1) the EC method, (2) the MLR method and (3) the full Neyman belt construction. The upper limits are shown as function of the number of expected background counts. In the large-sample limit, all methods give consistent results. In the limit of vanishing background counts, the EC method yields slightly stronger projected limits than the Neyman belt construction (consistent with Fig. 2 and the above discussion). In the intermediate regime, the Neyman belt construction shows a step-like structure, which is related to the discreteness of the Poisson likelihood. This is not visible in the EC results, but is a small effect almost everywhere. The MLR method yields results consistent with the Neyman belt construction. Deviations are due to MC noise.

Finally, for models 2 and 3, we observe relevant differences between the methods. The EC limits are here weaker ('more conservative') than the limits from the MLR method by up to $\sim 40\%$. We found that this is a rather general behaviour in the Poisson regime, which we observed for a large range of non-trivial scenarios. However, as discussed above, these large deviations are not observed in the single-bin case model 1.

To shed further light on the difference between the single-bin and multi-bin scenarios in the Poisson regime, we show in Fig. 6 for model 2 the expected exclusion limits derived from the EC and the MLR methods, as function of the normalization of the background component, θ_2. Three regimes can be clearly discriminated. For $\theta_2 \ll 10^{-5}$, the number of expected background counts over the entire range of x is negligible. This case is equivalent to a simple single-bin scenario, where the EC and MLR methods agree well (see also Fig. 5 for model 1). On the other hand, for values of $\theta_2 \gg 10^2$, we enter the Gaussian regime where the number of background counts is large over the entire range of x. Again, EC and MLR results agree well.

However, in the intermediate range, $\theta_2 \sim 10^{-3}$–10^{-1}, we find a plateau where the MLR and EC methods yield different results. In the plateau region, the number of background events is large at $2 < x < 10$, but negligible at $0 < x < 2$. This effectively reduces the
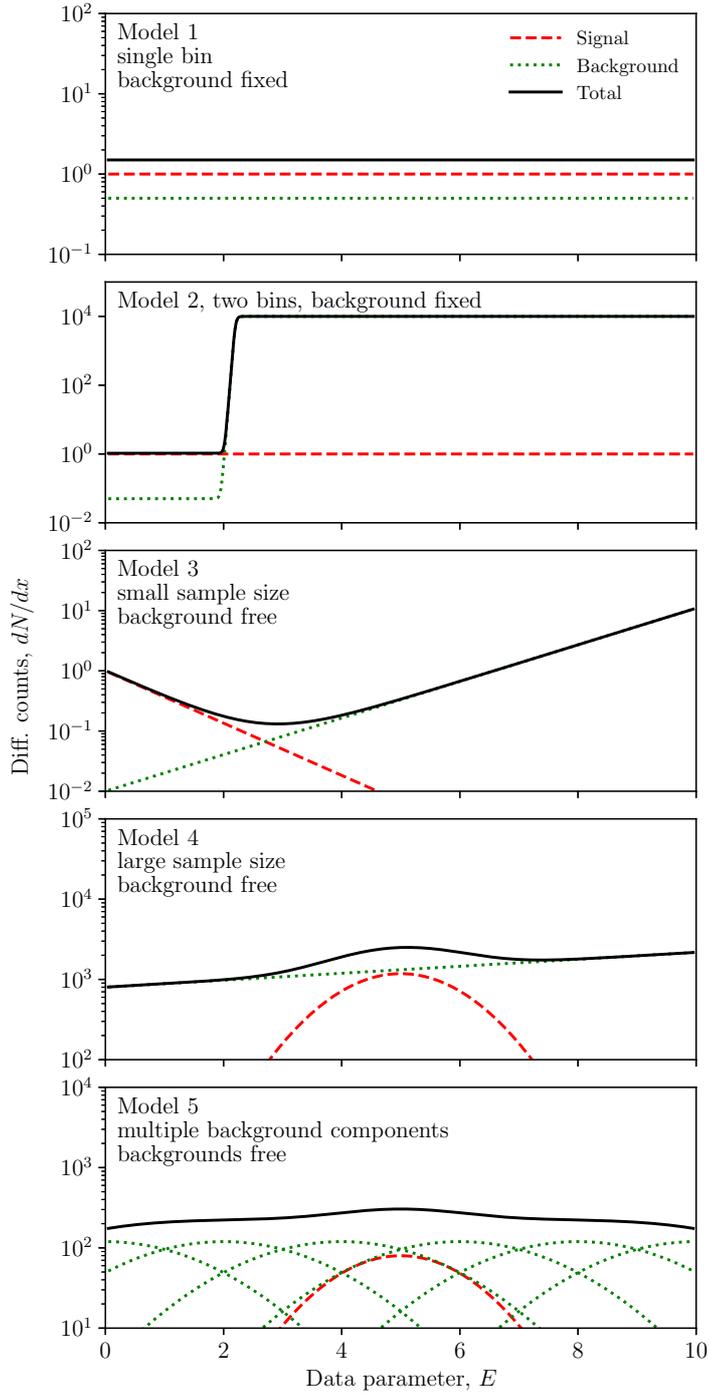
Figure 3. Five example models that we use to test the EC method against the full MLR method. The signal component refers to $\Psi_1(E)$, the background component to $\Psi_2(E)$. We also indicate whether background components are treated as fixed or free in the analysis. Model 5 has six background components, $\Psi_{2-7}(E)$, instead of one.

number of statistically relevant signal events by a factor of five, and should consequently weaken the expected exclusion limits by the same factor w.r.t. the $\theta_2 \ll 10^{-5}$ case. This
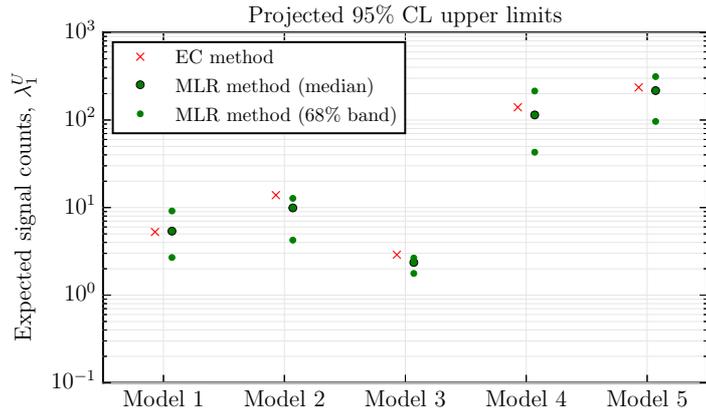
Figure 4. Expected exclusion limits for the five models in Fig. 3, derived using the ECM, Eq. (3.3), compared to the median limit from the full coverage-corrected MLR method. For comparison, we also show the 68% containment regions of the MLR limits. These regions contain, for multiple realizations of background-only data, 68% of the corresponding upper limits.
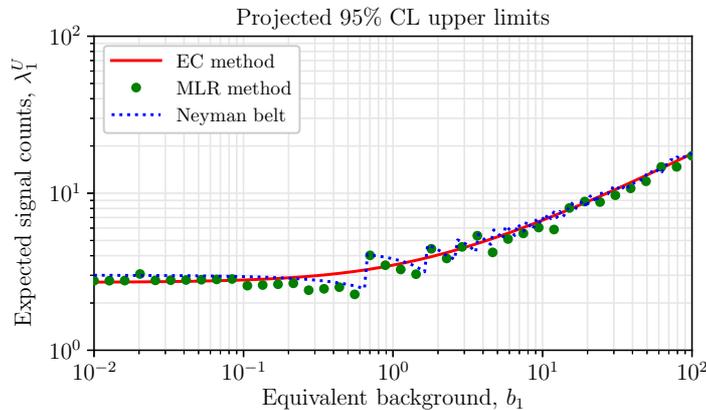
Figure 5. Various projected upper limits for model 1 from Fig. 3, as a function of the total equivalent background counts b_1: Limits from the EC method (red solid line). The median limit based on a full coverage-corrected MLR method (green dots). The median limit from the Neyman Belt construction (blue dotted line).

happens indeed for the limits derived using the EC method, but not for the limits from the MLR method. We trace this behaviour back to the fact that in the plateau region, the events from $2 < x < 10$ introduce a noise in the otherwise Poissonian likelihood from $0 < x < 2$. This effectively removes the discreteness of the Poisson distribution. As a consequence, the MLR limits in the plateau region can cover exactly, while the MLR limits in the $\theta_2 \ll 10^{-5}$ actually over-cover.

Scenarios like model 3 have a low- and a high-background regime, and hence behave similar to model 2 in the plateau region, leading to discrepancies between the MLR and the EC results (see Fig. 4). We never found the effect to exceed 40%.
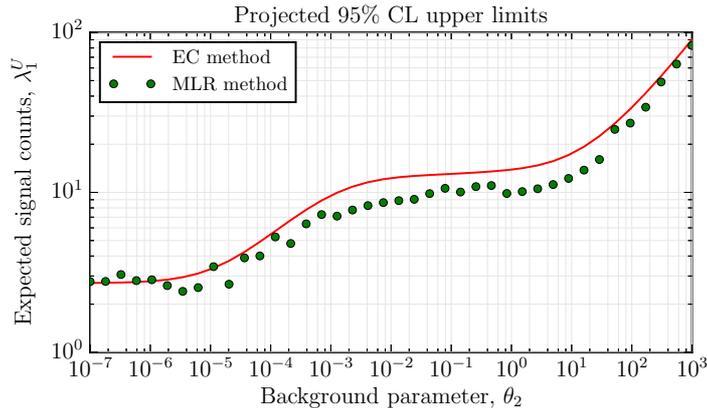
Figure 6. Projected upper limits for model 2 from Fig. 3, using the EC method and the coverage-corrected MLR method. See text for a detailed discussion.

3.2 Expected discovery reach

Equivalent counts method. The discovery limit for θ_1, *i.e.* the value of θ_1 that leads in 50% of the cases to a rejection of the $\theta_1 = 0$ hypothesis with a significance level α, can be approximately obtained by numerically solving the following equation for θ_1^D:

$$\left(s_1(\boldsymbol{\theta}^D) + b_1(\boldsymbol{\theta}^D)\right) \ln\left(\frac{s_1(\boldsymbol{\theta}^D) + b_1(\boldsymbol{\theta}^D)}{b_1(\boldsymbol{\theta}^D)}\right) - s_1(\boldsymbol{\theta}^D) = \frac{Z(\alpha)^2}{2} . \tag{3.6}$$

Here, we use the notation $\boldsymbol{\theta}^D = (\theta_1^D, \theta_2, \ldots, \theta_n)^T$, and s_1 and b_1 refer to the equivalent signal and background counts discussed at the end of Sec. 2.4. Heuristically, Eq. (3.6) is motivated by the analytic structure of profile likelihood ratios, see for instance discussion in Ref. [3]. However, its main motivation comes from the fact that it leads to approximately correct results both in the signal- and background- limited regimes.

It is useful to consider limiting cases. In the *background-limited case*, $b_1(\boldsymbol{\theta}^D) \gg s_1(\boldsymbol{\theta}^D)$, Eq. (3.6) implies that

$$\theta_1^D \simeq Z(\alpha) \cdot \sigma_1(\boldsymbol{\theta}) , \tag{3.7}$$

with $\boldsymbol{\theta}_0 = (0, \theta_2, \ldots, \theta_n)^T$. This is exactly what we expect for Gaussian background noise with variance σ_1^2. On the other hand, in the *signal-limited case*, $b_1(\boldsymbol{\theta}^D) \ll s_1(\boldsymbol{\theta}^D)$, one can show that the solution to Eq. (3.6) satisfies the following equation (details can be found in Appendix C.1):

$$\frac{b_1(\boldsymbol{\theta}^D)^{s_1(\boldsymbol{\theta}^D)}}{\Gamma\left(s_1(\boldsymbol{\theta}^D) + 1\right)} = \alpha \cdot \sqrt{\frac{Z^2(\alpha)}{s_1(\boldsymbol{\theta}^D)}} . \tag{3.8}$$

Here, $\Gamma(\cdot)$ is the gamma function. This equation is very similar to the exact expression derived from the Poisson distribution using Asimov data [3] in the low-background limit. To see this, note that, given some background $b_1 \ll 1$ and zero signal, the probability to detect λ or more photons is approximately $(b_1)^\lambda / \Gamma(\lambda + 1)$. If λ were the discovery reach corresponding to significance level α, this expression should equal α. The difference between this exact and the above approximate expressions is hence the square-root on the right-hand side of Eq. (3.8). In practice, this turns out to be a small effect, as we will see below.
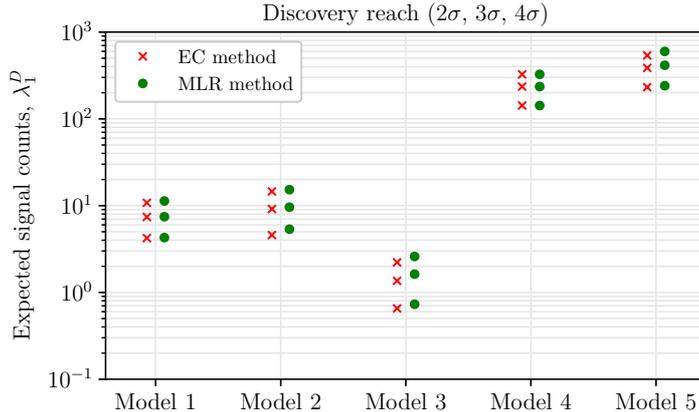
Figure 7. Expected discovery limits derived from the EC method, Eq. (3.6), compared wit full MLR results, for the five models in Fig. 3. Groups of three symbols show from bottom to top 2σ, 3σ and 4σ discovery limits (meaning that 50% of the measurements would lead to a detection with at least the indicated significance). Both methods yield consistent results, see detailed discussion in the text.

Comparison with exact methods. In Fig. 7 we compare the expected discovery limits derived using the EC method, Eq. (3.6), with the ones from the full MLR method, for the five example models in Fig. 3 (details can be found in Appendix B). We find that for all cases the EC results remain very close to the MLR results. Deviations are usually much less than 1σ, and largest for model 3 where the equivalent background is lowest. This good agreement is quite remarkable, given that some of the models are deeply in the Poisson regime.
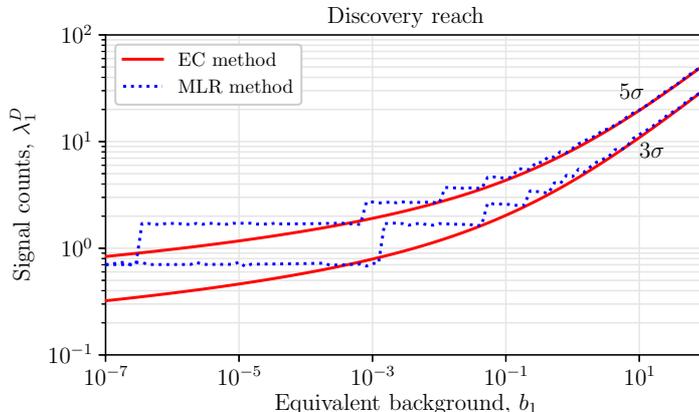
Figure 8. Comparison of 3σ (bottom) and 5σ (top) expected discovery limits based on the EC method, Eq. (3.6), and on the MLR method. The discreteness of the Poisson distribution becomes apparent for $b_1 \lesssim 1$, where the pure EC methods over-predicts the sensitivity by the indicated amount. See text for a detailed discussion.

To further investigate the limitations of our EC method in the low-background regime, we compare in Fig. 8 for model 1 the 3σ and 5σ expected discovery limits derived from the EC and the MLR methods. The approximate discovery limits resemble well the exact results down to equivalent backgrounds of one. For smaller equivalent backgrounds, the discreteness of the Poisson distribution starts to dominate the MLR method, which is not seen in the EC results. However, depending on the significance level, the agreement remains reasonably

good even down to and below $b_1 \sim 10^{-3}$. We note that, in very extreme cases, the EC method could lead to expected discovery limits that are much smaller than one. This can be prevented by the additional ad hoc requirement that the equivalent number of signal counts, $s_1(\boldsymbol{\theta})$, should be at least one.

3.3 Limitations of the Fisher approach

There are two important situation where the prescriptions for deriving expected exclusion limits, Eq. (3.3), and expected discovery limits, Eq. (3.6), will break down and give potentially wrong results.[4] Here we address both in detail.

Skewness of Poisson likelihood. The Fisher information matrix as defined in Eq. (2.1) encodes complete knowledge about the likelihood function, provided that the likelihood behaves like a multivariate normal distribution. This is equivalent to requiring that higher order derivatives in the expansion of the log-likelihood are negligible. Often, higher derivative terms will only cause deformations to Gaussian contours, but in the extreme case in which parameters are infinitely degenerate, the Fisher formalism will provide not only quantitatively but also qualitatively wrong results [6]. Non-Gaussianity is a common challenge in cosmology [24], but also relevant for the simple additive component model with the Poisson likelihood considered here.

As we saw above, using our EC method, one can obtain reasonable expected exclusion and discovery limits even in the deep Poisson regime, despite the fact that the likelihood functions are clearly non-Gaussian in that case. This is partially due to some lucky numerical coincides that happen to make our proposed prescriptions reasonably accurate. However, these mechanisms only apply to the signal-component of interest. The likelihood functions corresponding to the background (any non-signal) components should obey the usual Gaussianity constraints to ensure that the EC method can be applied.

In Appendix A.3, we study the behaviour of the Poisson likelihood function up to third order in the model parameters. Assuming that no strong degeneracies exist between parameters, one can derive a simple condition on the equivalent number of background events that should hold for all background components. It reads

$$b_i \gtrsim \frac{4t}{9f^2} \ , \tag{3.9}$$

where b_i is the number of *equivalent background* counts of component i, t the value of $\Delta(2 \ln \mathcal{L})$ at the boundary of the confidence region of interest, and f the tolerable fractional uncertainty of t at that boundary. If, for instance, we want to have 2σ intervals (for one dimension this corresponds to $t = 4$) with a fractional significance error of less than 20% (this corresponds to $f \simeq 40\%$, since t depends quadratically on the significance in standard deviations), this implies that the equivalent background for the components i should exceed $b_i \gtrsim 11$. In cases where parameter degeneracies might play a role, we recommend to use the full expressions provided in Appendix A.3 instead.

---

[4]This is the case for the additive component models considered in the present work. For more general Poisson problems, *e.g.* with shape uncertainties, the number of potential problems is larger.

Parameter degeneracies. When using the Fisher formalism, implicitly assumes that model parameters are unbound and only constrained by the likelihood function. This means that there is nothing that prevents part of the parameters θ_i to become negative. Consider as simple example a signal, $\Psi_i(E)$, that is exactly degenerate with a background component $k \neq i$, $\Psi_k(E) \propto \Psi_i(E)$. In that case, the Poisson part of the Fisher information matrix becomes singular, and the error of the signal component, σ_i, diverges. An arbitrarily large signal i can be compensated by an equally large negative background component k. Since this behaviour is usually unphysical (unless for instance absorption effects are part of the model), it must be prevented when performing Fisher forecasting.

A sufficient condition to exclude the problems with negative components reads

$$s \cdot \sigma_k(\boldsymbol{\theta}) < \theta_k \quad \text{for all} \quad k \neq i \,, \tag{3.10}$$

where s is the significance of interest in standard deviations. If errors are sufficiently small, the parameter θ_k will not cross zero. As will be discussed below in section 4, if the projected data alone is not sufficient to break the degeneracy between different background components or the background and the signal, it is possible (and necessary) to include additional constraints on parameters such that Eq. (3.10) is fulfilled. If this is not possible, the EC method cannot be directly applied.

4 Modeling of instrumental and background systematics

We give here a few instructive examples of how to model background uncertainties within the Fisher information framework.

4.1 Basic parameter systematics

In many cases of practical importance, additional information about nuisance parameters is available, which must be included in the sensitivity projections to obtain realistic results. Within a Bayesian approach, these additional constraints would be included as priors on the nuisance parameters. Within the Frequentist treatment, which we focus on here, a common approach is to include additional parameter constraints as effective likelihoods, as described in the following. These additional constraints can, for instance, come from 'sideband measurements' in signal-free regions of the data space.

If we assume that the constraints on parameter θ_i are Gaussian, with standard deviation ξ_i, the associated combined likelihood function can be written as

$$\mathcal{L}(\mathcal{D}|\boldsymbol{\theta}) = \mathcal{L}(\mathcal{D}|\boldsymbol{\theta})_{\text{pois}} \times \prod_i \mathcal{N}(\theta_i^A|\mu = \theta_i, \sigma^2 = \xi_i^2) \,. \tag{4.1}$$

Here, $\mathcal{N}$ refers to the PDF of a normal distribution with variance ξ_i^2 and Asimov value θ_i^A. It accounts for potential sideband measurements and similar external constraints. In this spirit, θ_i^A is taken to equal the mean value, $\theta_i^A = \theta_i$, when the average $\langle \cdot \rangle_{\mathcal{D}(\boldsymbol{\theta})}$ is applied in Eq. (2.1).

The resulting total Fisher information matrix can be split in a Poisson and a systematics part, $\mathcal{I}_{ij} = \mathcal{I}_{ij}^{\text{pois}} + \mathcal{I}_{ij}^{\text{syst}}$, where the latter is here diagonal and given by

$$\mathcal{I}_{ij}^{\text{syst}} = \delta_{ij} \frac{1}{\xi_i^2} \,. \tag{4.2}$$

The generalization to correlated systematic errors reads $\mathcal{I}_{ij}^{\text{syst}} = (\Sigma_{\text{syst}}^{-1})_{ij}$. We will show in a few examples how this is used in practice.

4.2 Example 1: Background systematics degenerate with the signal

We start with a simple example where we assume that some component of the background systematic is perfectly degenerate with the signal. More specifically, we consider a three-component model, where $\Psi_1(E)$ is the signal, $\Psi_2(E)$ the nominal background, and the component $\Psi_3(E) = \Psi_1(E)$ accounts for small positive or negative perturbations of this background. Note that this implies that $\mathcal{I}_{1i}^{\mathrm{pois}} = \mathcal{I}_{3i}^{\mathrm{pois}}$ for $i = 1, 2, 3$, which means that the Poisson part of the Fisher matrix is singular. We set the background normalization to $\theta_2 = 1$, and the mean background perturbation to zero, $\theta_3 = 0$. For the systematics part of the Fisher matrix, Eq. (4.2), we assume that the background perturbation θ_3 is externally constrained with a variance of $\xi_3^2 > 0$, and the background normalization θ_2 and the signal normalization are unconstrained, $\xi_1^2, \xi_2^2 \to \infty$.

In that case, one can show, by calculating the profiled Fisher information for component $i = 1$, that the variance of the signal component is given by (details can be found in Appendix C.2)

$$\sigma_1^2(\boldsymbol{\theta}) = (\sigma_1^{\mathrm{pois}})^2(\boldsymbol{\theta}) + \xi_3^2 \ . \tag{4.3}$$

This means that, as one might have expected, statistical and systematic errors are added in quadrature. We remark that constraints on the signal parameter have the opposite effect, $\sigma_1^{-2}(\boldsymbol{\theta}) = (\sigma_1^{\mathrm{pois}})^{-2}(\boldsymbol{\theta}) + \xi_1^{-2}$, and *decrease* the overall variance.

To illustrate Eq. (4.3), we consider expected exclusion limits on line-like signals on top of a power-law background plus an 'instrumental line'. We treat the power-law background as fixed, and assume that the instrumental line has identical width to the signal line. However, the normalization of the instrumental line is constrained by $\xi_3 = 1.0$. The fluxes of the background and signal components are illustrated in the upper panel of Fig. 9. Specifically we consider a Gaussian located at $E = 10$, with width $\sigma_E^2 = 0.2$. The power law has a slope of 1.4 and normalisation of three, $\theta_2 = 3.0$. The normalisation of the instrumental line is kept fixed with $\theta_3 = 1.0$.

The projected limits are shown in the lower panel of Fig. 9. When the signal becomes degenerate the with instrumental lines, the projected limits weaken, according to Eq. (4.3). We also recover the expected behaviour that when the normalization of the instrumental line is unconstrained (see Sec. 3.3) , the projected limit calculated by the Fisher information at the position of the instrumental line diverges.

4.3 Example 2: Background systematics described by correlation function

A more general approach towards the modeling of background systematics is to write the total flux as

$$\Phi(E) = \theta_1 \Psi_1(E) + (1 + \delta(E))\Psi_2(E) \ , \tag{4.4}$$

where again Ψ_1 denotes the signal, Ψ_2 the nominal background (we fix $\theta_2 = 1$ throughout), and $\delta(E)$ parameterizes small deviations from the background model. In general, systematic uncertainties in the background will be correlated as a function of energy, and one can define the covariance function

$$\langle \delta(E)\delta(E') \rangle = \Sigma_\delta(E, E') \ , \tag{4.5}$$

and we assume $\langle \delta(E) \rangle = 0$. More specifically, $\delta(E)$ can be thought of as *Gaussian random field* with mean zero, whose behaviour is completely determined by the covariance function. It incorporates information both about the variance and correlation of the systematic.
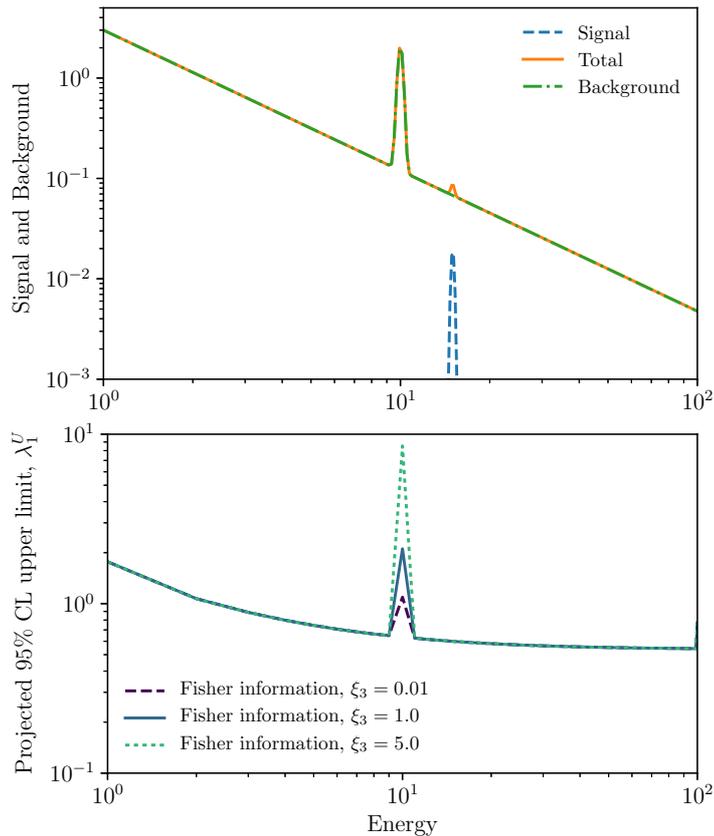
Figure 9. *Upper panel:* Example of a fixed power law background with one instrumental lines at $E = 10$. We show an example of a signal line with identical width to the instrumental lines (here at $E = 15$). *Lower panel:* We calculate expected exclusion limit using the EC method, as a function of the signal position. The limits becomes more constraining at higher energies, due to the increased SBR. However, there is loss of constraining power when the signal is degenerate with the instrumental line, depending on the uncertainty of the instrumental line, ξ_3.

For any practical calculations, one needs to discretize the field $\delta(E)$. One simple way of doing that is to write $\delta(E)$ as a step function,

$$\delta(E) = \sum_{i=1}^{N} \xi_i \chi_{\Delta E_i}(E) \,, \tag{4.6}$$

with the selector function

$$\chi_{\Delta E_i} = \begin{cases} 1 \text{ if } E \in \Delta E_i \\ 0 \text{ if } E \notin \Delta E_i \end{cases}, \tag{4.7}$$

where the (very small) energy bins ΔE_i cover the entire energy range of interest, and E_i will denote the corresponding mean of each energy bin. Furthermore, ξ_i are constrained by a multivariate normal distribution, with a covariance matrix defined by $\Sigma_\delta(E_i, E_j)$. The free parameters in the present example are then $\boldsymbol{\theta} = (\theta_1, \xi_1, \ldots, \xi_N)$.

One can now show that the resulting profiled Fisher information for the signal (assuming a sufficiently fine binning which captures all relevant structure of $\Psi_1(E)$), is given by (details

can be found in Appendix C.2)

$$\widetilde{\mathcal{I}}_{11} = \sum_{ij} \frac{\Psi_1}{\Psi_2}(E_i) D_{ij}^{-1} \frac{\Psi_1}{\Psi_2}(E_j) , \tag{4.8}$$

where we defined the combined covariance matrix

$$D_{ij} \equiv \frac{\delta_{ij}\Phi(E_i)}{\Delta E_i [\Psi_2(E_i)]^2} + \Sigma_\delta(E_i, E_j) , \tag{4.9}$$

which includes both the effects of Poisson noise and background systematics. This expression becomes more simple in the regime where $\Psi_1 \ll \Psi_2$ since then $\Phi(E) = \Psi_2(E)$. In the limit of no systematics, $\Sigma_\delta \to 0$, we recover the standard expression for the Fisher information of the signal component, Eq. (2.13). On the other hand, in the large sample limit, where Ψ_2 becomes large and hence the first term in Eq. (4.9) small, only the covariance matrix Σ_δ matters and determines the limiting accuracy at which Ψ_1 can be measured. Note that results are independent of the bin size as long as it is sufficiently small to fully resolve the discriminating aspects of the different model components.

This approach of estimating the effect of background uncertainties on expected experimental sensitivities has been already used by some of the present authors in Ref. [13]. We will provide another example below in Sec. 5, in context of the Fisher information flux.

5 Strategy optimization

Experimental design, or the planning of astronomical observational campaigns, often make use of the SNR of some signals of interest (for the simple additive component models, Eq. (2.10) that we discussed above, this corresponds to $\propto \Psi_i(E)/\sqrt{\Phi(E)}$). One of the goals is to maximize the exposure of energy and/or spatial regions that provide the largest SNR for some component i, which then leads to the tightest constraints on the model parameter θ_i.

We will show here that the above SNR is the simplest realization of the Fisher *information flux*, which we newly introduce here. However, the latter is much more general and can also naturally account for the non-local and saturation effects of background and instrumental systematics. We will illustrate this in an example that makes use of the treatment of correlated background systematics that we discussed in the previous section.

5.1 Fisher information flux

In this subsection we will look at the model spectrum Φ as function of the sky coordinate, $\Phi(\Omega|\boldsymbol{\theta})$, such that $I_i(\Omega)$ is the intensity of the signal or background component,[5] and $\mathcal{E}(\Omega)$ is the exposure of the instrument towards Ω, see Eq. (2.11).

With this, we can define the differential *Fisher information flux* that corresponds to parameter pair (i, j) as functional derivative w.r.t. the exposure at position Ω. It is given by

$$\frac{d\mathcal{F}_{ij}}{d\Omega}(\boldsymbol{\theta}, \mathcal{E}) \equiv \frac{\delta \mathcal{I}_{ij}(\boldsymbol{\theta}, \mathcal{E})}{\delta \mathcal{E}(\Omega)} , \tag{5.1}$$

---

[5]To keep the notation simple, we ignore here the effects of the instrument point-spread function or energy dispersion. They can be added straightforwardly.

where we made explicit that the Fisher information is in general a function of the exposure map $\mathcal{E}(\Omega)$. If we consider the Poisson part of the Fisher information alone, we find

$$\frac{\delta \mathcal{I}_{ij}^{\mathrm{pois}}(\boldsymbol{\theta}, \mathcal{E})}{\delta \mathcal{E}(\Omega)} = \frac{1}{\theta_i \theta_j} \frac{I_i(\Omega) I_j(\Omega)}{I(\Omega|\boldsymbol{\theta})} \ . \tag{5.2}$$

The diagonal part of the Fisher information flux corresponds here to the SNR of component i, and the non-diagonal parts provide information about the degeneracy of the components pairs (i, j).

We emphasize that the right-hand side of Eq. (5.2) does *not* depend on the exposure anymore, and is hence constant during the course of the measurement. Similarly, external constraints like in Eq. (4.2) do not depend on the exposure. In these cases, the full Fisher information flux in Eq. (5.1) would equal the Fisher information flux of the Poisson likelihood, Eq. (5.2). This simple situation changes drastically when considering the *effective information flux* for a subset of the model parameters, as we will see in the next subsection.

The information gain about the parameter pair (i, j) that is obtained by increasing the exposure towards direction Ω by the infinitesimal amount $\delta \mathcal{E}(\Omega)$ is given by

$$\delta \mathcal{I}_{ij}(\boldsymbol{\theta}, \mathcal{E}) = \int d\Omega \, \delta \mathcal{E}(\Omega) \frac{d\mathcal{F}_{ij}}{d\Omega}(\boldsymbol{\theta}, \mathcal{E}) \ . \tag{5.3}$$

As a simple application, let us assume that the change in exposure per time is given by

$$\frac{d\mathcal{E}(\Omega)}{dt} = A_{\mathrm{eff}} \frac{dT_{\mathrm{obs}}(\Omega)}{dt} \ , \tag{5.4}$$

where we factored out the effective area, A_{eff}, and $T_{\mathrm{obs}}(\Omega)$ is the accumulated observation time in direction Ω. Then, the information gain per unit time is given by

$$\frac{d\mathcal{I}_{ij}(\boldsymbol{\theta}, \mathcal{E})}{dt} = A_{\mathrm{eff}} \int d\Omega \, \frac{dT_{\mathrm{obs}}(\Omega)}{dt} \frac{d\mathcal{F}_{ij}}{d\Omega}(\boldsymbol{\theta}, \mathcal{E}) \ . \tag{5.5}$$

Integrating this over time would again give the total information obtained in the observation. Note that these equations remain valid also for the effective information flux that we discuss in the next subsection.

5.2 Effective information flux

In order to quantify information gain about PoIs in presence of background and other uncertainties, the above concept of Fisher information flux needs to be extended to the profiled Fisher information that we introduced in Eq. (2.9). This can be done straightforwardly by applying the functional derivative with respect to $\mathcal{E}(\Omega)$, Eq. (5.1), to the profiled Fisher information.

The full expression for the *effective information flux* for the PoI $(\theta_1, \ldots, \theta_k)$ reads

$$\frac{d\widetilde{\mathcal{F}}_A}{d\Omega} = \frac{d\mathcal{F}_A}{d\Omega} - \frac{d\mathcal{F}_C^T}{d\Omega} \mathcal{I}_B^{-1} \mathcal{I}_C + \mathcal{I}_C^T \mathcal{I}_B^{-1} \frac{d\mathcal{F}_B}{d\Omega} \mathcal{I}_B^{-1} \mathcal{I}_C - \mathcal{I}_C^T \mathcal{I}_B^{-1} \frac{d\mathcal{F}_C}{d\Omega} \ , \tag{5.6}$$

where as above A refers to the $k \times k$ part of the Fisher information matrix that corresponds to the PoI, B refers to the $(n - k) \times (n - k)$ part for the nuisance parameters, and C to the mixing between nuisance parameters and the PoI. This expression appears lengthy, but

straightforward to evaluate analytically or numerically if the Fisher information matrix and Fisher information flux are already known.

The effective information flux has a number of surprising and useful properties. First, it is in general not constant in time (in contrast to the plain information flux in Eq. (5.1)). In the examples considered here, this is due to *saturation effects*, which are related to the observation reaching the systematic limited regime. Second, it is *non-local*, in the sense that it for instance depends on the past (non-)observation of sidebands that could help to characterize the backgrounds in the signal region. Technically, the non-locality due to the fact that the full Fisher information matrix appears in Eq. (5.6), which includes integrals of the signal and background intensities over Ω. We will illustrate these two aspects in the following final example.

Saturation effects. We demonstrate the saturation effects of the effective information flux with a non-trivial examples from Sec. 4.3. There, we discussed how to treat correlated background systematics around a fixed background in the Fisher formalism, see Eq. (4.4).

For definiteness, we consider a signal component that consists of two Gaussian peaks,

$$I_1(E) = 0.05 \times \mathcal{N}(E|\mu = 2, \sigma^2 = 0.01) + \mathcal{N}(E|\mu = 6, \sigma^2 = 4) , \tag{5.7}$$

one of which is narrow and the other wide. The background is taken to be flat, $I_2(E) = 1$. The corresponding model count spectra $\Psi_{1,2}(E)$ are obtained by multiplication with the exposure $\mathcal{E}(E)$, as in Eq. (2.11). Lastly, we define the covariance function of the background deviations as,

$$\Sigma_\delta(E, E') = 0.01 \times \mathcal{N}(E|\mu = E', \sigma^2 = 1) + 0.01 \times \mathcal{N}(E|\mu = E', \sigma^2 = 2) . \tag{5.8}$$

For simplicity, we focus on the case $\theta_1 = 1$. A sum of two Gaussians was chosen to demonstrate the techniques ability to account for multiple correlation lengths.

The profiled Fisher information for the signal component $\Psi_1(E)$ is given by Eq. (4.8) above. The corresponding effective differential information flux at energy E_k can be obtained by differentiating this expression w.r.t. exposure in energy bin ΔE_k (remember that we have bins in energy for practical purposes). The resulting *effective information flux* of the signal reads

$$\frac{d\widetilde{\mathcal{F}}_{11}}{dE}(E_k) = \sum_{ij} \frac{I_1}{I_2}(E_i) D_{ik}^{-1} \frac{I(E_k)}{\Delta E_k^2 \mathcal{E}(E_k)^2 I_2^2(E_k)} D_{kj}^{-1} \frac{I_1}{I_2}(E_j) . \tag{5.9}$$

Here, D_{ij} refers to the combined covariance matrix defined in Eq. (4.9).

It is instructive to consider two limiting cases. If background uncertainties are negligible w.r.t. Poisson noise, the first term in the right-hand side of Eq. (4.9) dominates, and we obtain the pure Poisson information flux

$$\frac{d\widetilde{\mathcal{F}}_{11}}{dE} = \frac{I_1(E)^2}{I(E)} . \tag{5.10}$$

On the other hand, for a sufficiently large exposure, the second term in the right-hand side of Eq. (4.9) can dominate, and D_{ij} becomes constant in time. In that case, the effective information flux scales like $\propto \mathcal{E}^{-2}$, leading to a finite total measured information even for very large integration times.[6]
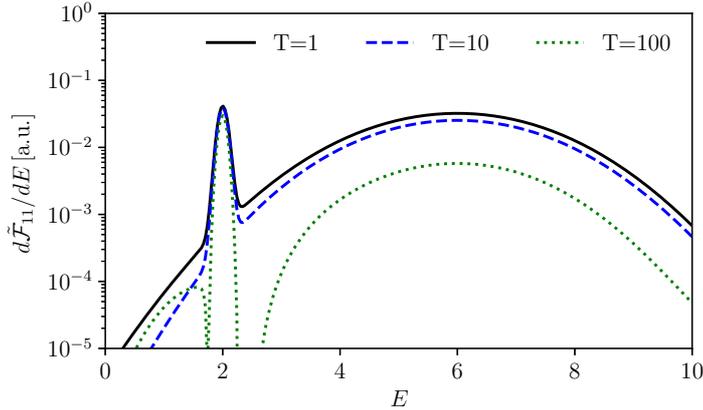
Figure 10. Example for the saturation effects of the effective information flux. We show here the effective information flux for a signal with narrow and a broad spectral component, on top of a background with correlated uncertainties, after different observation times $T_{\rm obs}$. See Eq. (5.9) and text for further details.

In Fig. 10 we show the effective information flux from Eq. (5.9), for different values of the past observation time $T_{\rm obs}$ (remember that $\mathcal{E}(E) = A_{\rm eff}T_{\rm obs}$, and we set $A_{\rm eff} = 1$). For small observation times, the flux essentially corresponds to the pure Poisson contribution in Eq. (5.10). However, for larger observation times, the information flux from the broad peak around $E = 6$ becomes increasingly suppressed. This is due to the fact that this broad feature is significantly degenerate with the modeled background uncertainties. On the other hand, the flux from the narrow signal component around $E = 2$, which has a width that is smaller than the correlation scale of the modeled background systematics, remains practically constant.
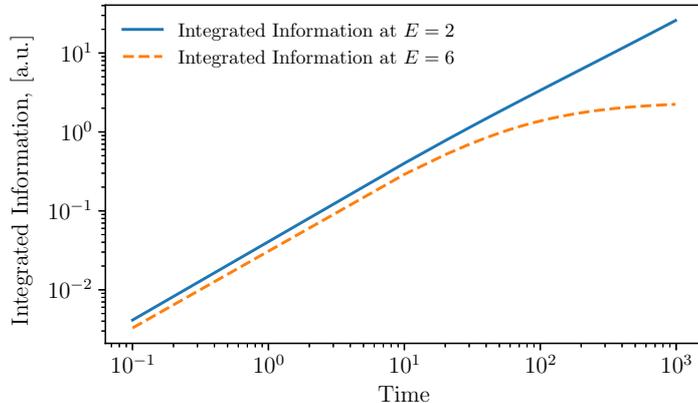
Figure 11. The cumulative information flux at two reference energies, corresponding to the scenario in Fig. 10, but integrated over the observation time, as function of the total observation time $T_{\rm obs}$.

---

[6]This argument does not hold if the vector $x_i \equiv I_1/I_2(E_i)$ has components with zero eigenvalues w.r.t. the matrix $M_{ij} \equiv \Sigma_\delta(E_i, E_j)$. Components with zero eigenvalues correspond to characteristics of the signal that are completely uncorrelated with the modeled background uncertainties, and give rise to non-zero contributions to the effective information flux even after very large integration times. This often undesired behaviour can be removed by adding a small diagonal contribution to the background uncertainty M, which also improves the numerical stability of the matrix inversions.

To further illustrate the saturation effects when measuring over a long time, we show in Fig. 11 the cumulative information obtained by integrating the effective information flux from Eq. (5.9) at the peaks of the two features, $E = 2$ and $E = 6$. It is clear that at $T = 10^2$ the information obtained from observing the broad peak becomes saturated, while the sharper peak continues to provide information.

Non-locality. We demonstrate the non-locality of the effective information flux with a simple two-component example. With non-locality, we mean that the information flux at E depends in general on the past observation history of $E' \neq E$. This makes sense, since usually a comparable exposure of different observational regions is required to break degeneracies between various model components.
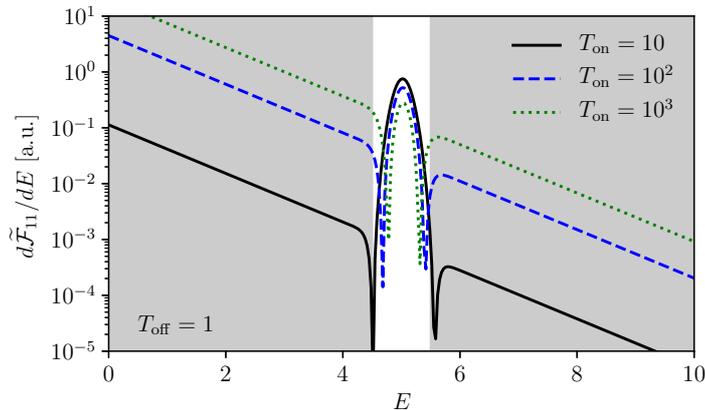
Figure 12. Example for the non-locality of the effective information flux. We show the effective information flux for a narrow signal on top an exponential background with free normalization. After an initial observation of the sidebands for $T_{\rm off} = 1$, observations are assumed to only take place in the narrow range $E = 4.5$–5.5, and we show the effective information flux after different signal observation times $T_{\rm on}$. See text for details.

In our simple two-component example, the signal flux, $I_1(E)$, is given by a normal distribution, centered at $E = 5$, with a width of $\sigma = 0.2$. The background flux is given by $I_2(E) = 5 \exp(-(E - 5))$. We assume that after an initial observation of the sidebands with $T_{\rm off} = 1$, observations only take place in the narrow range $E = 4.5$–5.5. We show the resulting effective information flux in Fig. 12. At early observation times $T_{\rm on} = 10$, the information flux is completely dominated by observations of the signal region, since due to the initial sideband observations the background in the signal region is already reasonably constrained. However, with growing observation time, the non-observations of the sidebands, becomes increasingly problematic. Consequently, the information flux of the sidebands grows continuously.

6 Conclusions

We introduced new methods and concepts for efficient and informative forecasting in astroparticle physics and DM searches, based on the powerful Fisher matrix formalism and unbinned Poisson likelihoods. The Fisher information matrix, Eq. (2.1), is a way of quantifying the maximum information that an observation carries about a set of model parameters. It

is at the core of many more advanced statistical methods, and heavily used in other branches of science.

We introduced compact expressions for the approximate derivation of expected exclusion and discovery limits, Eqs. (3.3) and (3.6) (*equivalent counts method*). The equivalent counts method is based on new definitions for the equivalent number of signal and background events, Eqs. (2.18) and (2.19). These are solely based on the Fisher information matrix. The equivalent counts method leads to surprisingly accurate results, even deeply in the Poisson regime, as we showed by comparison with the exact Neyman belt construction and maximum likelihood ratio techniques. In this work, we focused on additive component models and Poisson processes, and neglected the effects of shape uncertainties.

We furthermore showed in two examples how systematic uncertainties can be efficiently accounted for within the Fisher formalism. In the first example, we assumed that a component of the background is completely degenerate with the shape of the signal, Eq. (4.3). In the second example, we modeled background uncertainties using the idea of Gaussian random fields with arbitrary correlation functions. The resulting expressions, Eq. (4.8), can be efficiently solved numerically by matrix inversion, without profiling or marginalizing over the potentially thousands of nuisance parameters.

Finally, we introduced the new concept of *information flux*, which we obtained from the Fisher information matrix by applying a functional derivative w.r.t. the instrument exposure, Eq. (5.1). For unbinned Poisson likelihoods, it is equivalent to the commonly used signal-to-noise ratio. However, when including the effect of nuisance parameters, the resulting *effective information flux* accounts automatically for the non-local properties and saturation effects of background and instrumental uncertainties. We illustrated these effects in two examples in Figs. 10 and 12.

Our motivation for this work was to provide a both solid and efficient statistical framework for the systematic study of optimal search strategies for a large range of dark matter models in indirect and other searches which will be the subject of future publications. We furthermore plan to expand the discussion towards models with shape uncertainties, model discrimination and Fisher information geometry.

In summary, we showed how to make the powerful Fisher matrix formalism useful for typical problems in astroparticle physics and DM searches. The equivalent counts method for calculating expected exclusion and discovery limits is applicable in a large range of diverse situations, ranging from direct DM searches to the detectability of extended gamma-ray sources. The effective Fisher information flux is a flexible tool for search strategy optimization, and we expect it to be particularly interesting when confronted with a large number of potential targets, like in indirect searches for DM or a large number of analysis channels in particle colliders. In this work, we just scratched the surface of what can be done with the Fisher formalism, and expect fruitful further theoretical developments of the formalism in the future.

Acknowledgments

References

[1] G. Cowan, *Statistical data analysis* (1998).

[2] F. James, *Statistical methods in experimental physics* (2006).

[3] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, Eur. Phys. J. **C71**, 1554 (2011), [Erratum: Eur. Phys. J.C73,2501(2013)], arXiv:1007.1727 [physics.data-an] .

[4] S. S. Wilks, The Annals of Mathematical Statistics **9**, 60 (1938).

[5] H. Chernoff, The Annals of Mathematical Statistics **25**, 573 (1954).

[6] L. Wolz, M. Kilbinger, J. Weller, and T. Giannantonio, JCAP **1209**, 009 (2012), arXiv:1205.3984 [astro-ph.CO] .

[7] M. Vallisneri, Phys. Rev. **D77**, 042001 (2008), arXiv:gr-qc/0703086 [GR-QC] .

[8] E. Sellentin, M. Quartin, and L. Amendola, Mon. Not. Roy. Astron. Soc. **441**, 1831 (2014), arXiv:1401.6892 [astro-ph.CO] .

[9] A. Heavens, (2016), 10.3390/e18060236, arXiv:1606.06455 [astro-ph.CO] .

[10] J. Brehmer, K. Cranmer, F. Kling, and T. Plehn, (2016), arXiv:1612.05261 [hep-ph] .

[11] S. Camera, M. Fornasa, N. Fornengo, and M. Regis, JCAP **1506**, 029 (2015), arXiv:1411.4651 [astro-ph.CO] .

[12] T. Bringmann, A. Galea, A. Hryczuk, and C. Weniger, (2016), arXiv:1610.04613 [hep-ph] .

[13] R. Bartels, D. Gaggero, and C. Weniger, (2017), arXiv:1703.02546 [astro-ph.HE] .

[14] E. Charles *et al.* (Fermi-LAT), Phys. Rept. **636**, 1 (2016), arXiv:1605.02016 [astro-ph.HE] .

[15] A. A. Abdo *et al.* (Fermi-LAT), Astrophys. J. Suppl. **208**, 17 (2013), arXiv:1305.4385 [astro-ph.HE] .

[16] A. Albert, G. A. Gomez-Vargas, M. Grefe, C. Munoz, C. Weniger, E. D. Bloom, E. Charles, M. N. Mazziotta, and A. Morselli (Fermi-LAT), JCAP **1410**, 023 (2014), arXiv:1406.3430 [astro-ph.HE] .

[17] G. Casella and R. L. Berger, *Statistical Inference* (2002).

[18] C. Radhakrishna Rao, *Information and the accuracy attainable in the estimation of statistical parameters*, Vol. 37 (1945) pp. 81–91.

[19] H. Cramér, *Mathematical Methods of Statistics (PMS-9)*, Vol. 9 (Princeton university press, 2016).

[20] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic geometry and its applications* (John Wiley & Sons, 2013).

[21] J.-F. Coeurjolly and J. Mller, 10.3150/13-BEJ516, 1407.0249v1 .

[22] J. Neyman, Phil. Trans. Roy. Soc. Lond. **A236**, 333 (1937).

[23] G. J. Feldman and R. D. Cousins, Phys. Rev. **D57**, 3873 (1998), arXiv:physics/9711021 [physics.data-an] .

[24] E. Sellentin, M. Quartin, and L. Amendola, Mon. Not. Roy. Astron. Soc. **441**, 1831 (2014), arXiv:1401.6892 [astro-ph.CO] .

[25] W. A. Rolke, A. M. Lopez, and J. Conrad, Nucl. Instrum. Meth. **A551**, 493 (2005), arXiv:physics/0403059 [physics] .

A Poisson likelihood properties

We discuss here the general Poisson likelihood function that we use in the main part of the paper, its higher-order derivatives, expectation values and skewness.

A.1 Generalized Poisson likelihood

We concentrate in this section on univariate data (the generalization to the multivariate case is straightforward). A typical example are photon energy spectra, measured over some fixed energy range, $E^- \dots E^+$. The data is then fully described by an unordered list of the energies of the N measured photons,

$$\mathcal{D} \equiv \{E_1, E_2, \dots, E_N\} \, . \tag{A.1}$$

The corresponding PDF is

$$P(\mathcal{D}|\boldsymbol{\theta}) = \frac{e^{-\mu(\boldsymbol{\theta})}}{N!} \prod_{i=1}^{N} \Phi(E_i|\boldsymbol{\theta}) \, , \tag{A.2}$$

where $\Phi(E|\boldsymbol{\theta})$ is the model counts spectrum of the photons, and the expected total number of events can be calculated as

$$\mu(\boldsymbol{\theta}) \equiv \int_{E^-}^{E^+} dE \, \Phi(E|\boldsymbol{\theta}) \, . \tag{A.3}$$

The PDF in Eq. (A.2) is correctly normalized to one, which can be checked integrating over photon energies and summing over N.

The unbinned count map $\mathcal{C}(E)$, introduced in Eq. (2.4), carries exactly the same information as the unordered photon list $\mathcal{D}$. It turns out to be useful to rewrite the Poisson likelihood as function of the unbinned count map $\mathcal{C}(E)$. A formal expression that is completely equivalent to Eq. (A.2) is

$$P(\mathcal{C}|\boldsymbol{\theta}) = \exp\left(\int dE \left[\mathcal{C} \ln \Phi - \Phi \right] - \Gamma(N+1) \right) \, , \tag{A.4}$$

where we used the gamma function instead of the factorial, $N! = \Gamma(N+1)$, and the total number of measured events is given by

$$N \equiv \int_{E^-}^{E^+} dE \, \mathcal{C}(E) \, . \tag{A.5}$$

The key advantage is here that the domain on which $P(\mathcal{C}|\boldsymbol{\theta})$ is defined can be immediately extended to arbitrary functions $\mathcal{C}(E)$, which can be continuous and/or feature non-integer total measured events numbers. This will become useful below.

The likelihood function corresponding to the Poisson process is given by

$$\mathcal{L}(\mathcal{C}|\boldsymbol{\theta}) \equiv A(\mathcal{C})P(\mathcal{C}|\boldsymbol{\theta}) \, , \tag{A.6}$$

where $A(\mathcal{C})$ is an arbitrary positive function of the data that does not affect the discussion. The logarithm of Eq. (A.6), together with the choice $A(\mathcal{C}) = \exp(-\Gamma(N+1))$, leads Eq. (2.3) in the main text.

A.2 Expectation values and 'Asimov data'

Writing the Poisson log-likelihood in the form of Eq. 2.3 has the advantage that it is a linear function of the counts map $\mathcal{C}$. This means that averages over projected data are trivial. Remember that the unbinned count map averaged over many realizations of model $\boldsymbol{\theta}$ just equals the expected count map, $\langle\mathcal{C}\rangle_{\mathcal{D}(\boldsymbol{\theta})} = \Phi(\boldsymbol{\theta})$. Hence, for linear functions of $\mathcal{C}$, averaging over data is equivalent to substituting the unbinned counts map by the 'Asimov' data set [3], $\mathcal{C} \to \Phi(\boldsymbol{\theta})$.

Maximizing $\ln\mathcal{L}$ with respect to $\boldsymbol{\theta}$ requires that $\partial\ln\mathcal{L}/\partial\theta_k = 0$, which is equivalent to

$$\int_{E^-}^{E^+} dE \left(\mathcal{C}(E)\frac{\Psi_k(E)}{\Phi(E|\boldsymbol{\theta})} - \Psi_k(E) \right) = 0 \ , \tag{A.7}$$

for all components k. We adopted here the additive component model defined in Eq. (2.10). Replacing $\mathcal{C}$ by Asimov data for model $\boldsymbol{\theta}$ yields zero, as expected.

For the additive component model, averages over higher order derivatives of the log-likelihood take a simple form. The n-th order derivative (for $n \geq 2$) reads

$$\left\langle \frac{\partial^n(-\ln\mathcal{L})}{\partial\theta_{k_1}\ldots\partial\theta_{k_n}} \right\rangle_{\mathcal{D}(\boldsymbol{\theta})} = (-1)^n(n-1)! \int_{E^-}^{E^+} dE \ \frac{\Psi_{k_1}(E)\ldots\Psi_{k_n}(E)}{\Phi^{n-1}(E|\boldsymbol{\theta})} \ . \tag{A.8}$$

For $n = 2$, we recover Eq. (2.13) in the main text. The expression for $n = 3$ is important for the discussion of the non-Gaussianity effects in the next subsection.

A.3 The Gaussian regime

We start by Taylor expanding the log-likelihood around the model parameter $\boldsymbol{\theta}$,

$$\ln\mathcal{L} = \text{const} + \sum_i \Delta\theta_i\frac{\partial\ln\mathcal{L}}{\partial\theta_i} + \frac{1}{2!}\sum_{ij}\Delta\theta_i\Delta\theta_j\frac{\partial^2\ln\mathcal{L}}{\partial\theta_i\partial\theta_j} + \frac{1}{3!}\sum_{ijk}\Delta\theta_i\Delta\theta_j\Delta\theta_k\frac{\partial^3\ln\mathcal{L}}{\partial\theta_i\partial\theta_j\partial\theta_k} + \ldots \ , \tag{A.9}$$

where $\Delta\boldsymbol{\theta}$ is the deviation from the expansion point. If we average the $\ln\mathcal{L}$ now over model realizations $\mathcal{D}(\boldsymbol{\theta})$, and assume Poisson likelihoods, this expansion can be written as

$$\langle -\ln\mathcal{L}(\boldsymbol{\theta}+\Delta\boldsymbol{\theta})\rangle_{\mathcal{D}(\boldsymbol{\theta})} = \text{const} + \frac{1}{2}\sum_{ij}\Delta\theta_i\Delta\theta_j\left(\mathcal{I}_{ij} + \frac{2}{3}\sum_k\Delta\theta_k\frac{\partial\mathcal{I}_{ij}}{\partial\theta_k}\right) + \ldots \ . \tag{A.10}$$

Note that the additional factor two in front of $\partial\mathcal{I}_{ij}/\partial\theta_k$ comes from the fact that the derivative is here also affecting the parameters of the Asimov data. The precise factor depends on the actual likelihood function and its dependence on the data.

In order to understand the impact of higher-order terms in Eq. (A.10), it is convenient to think about how they affect the significance at the boundaries of confidence intervals. Naively, the boundary of a confidence interval that extends to a threshold value t in a certain direction would correspond to the ellipsoid constructed by values of $\Delta\theta_i$ that satisfy the equation

$$\sum_{ij}\Delta\theta_i\Delta\theta_j\mathcal{I}_{ij} = t \ . \tag{A.11}$$

When taking into account third-order terms from Eq. (A.10), the *actually* realized threshold value at point $\Delta\theta_i$ on the ellipsoid changes by

$$\Delta t = \frac{2}{3} \sum_{ijk} \Delta\theta_i \Delta\theta_j \Delta\theta_k \frac{\partial \mathcal{I}_{ij}}{\partial \theta_k} \ . \tag{A.12}$$

If $\Delta t \ll t$ at all points of the ellipsoid, higher-order (more precisely third order) terms can be indeed ignored and do not affect the result.

In practice, it is usually sufficient to concentrate on the principal axes of the ellipsoid, which correspond to the eigenvectors of $\mathcal{I}_{ij}$. For simplicity, we will here further assume that there are no strong degeneracies between the components. In that case, the principal axes approximately align with the individual parameters θ_i. We can then, for every direction i, require that

$$\frac{\Delta t}{t} = \frac{2}{3} \frac{\Delta\theta_i}{\mathcal{I}_{ii}} \frac{\partial \mathcal{I}_{ii}}{\partial \theta_i} < f \ , \tag{A.13}$$

which means that the fractional change in the threshold value t should be smaller than f. For the Poisson likelihood that we assumed already above in Eq. (A.10), we can use $\Delta\theta_i (\mathcal{I}_{ii})^{-1} \partial \mathcal{I}_{ii}/\partial\theta_i = \Delta\theta_i \sqrt{\mathcal{I}_{ii}}/\sqrt{b_i} = \sqrt{t/b_i}$, where we used the definition of the equivalent background in Eq. (2.19). This implies the condition

$$b_i > \frac{4t}{9f^2} \ , \tag{A.14}$$

which is identical to Eq. (3.9) in the main text and further discussed there.

We emphasize that this is a quite naive estimate, and does not take into account the possible effects of parameter degeneracies, deviations of the log-likelihood ratio from a chi-square distribution, etc. But it gives a useful heuristics for when the Fisher formalism is safe to use, and when it should be used with care. As we have seen in Sec. 3, if the appropriate prescriptions are used, reasonable results for expected exclusion and discovery limits on *one* signal component of interest can be obtained even for vanishing equivalent backgrounds, as long as the other components are well constrained and behaved.

B Expected exclusion and discovery limits

We describe here very briefly some of the exact methods that are used to calculate expected exclusion and discovery limits. These are in the main text compared against the results from our EC method.

B.1 Neyman belt construction

For details about the Neyman belt construction, we refer to Ref. [23]. We repeat here just the technical result. A conventional one-sided upper limit on the number of expected signal events, s, given k observed events and b expected background events, is given by the s^U that satisfies the equation

$$\sum_{k' \leq k} P(k'|s^U + b) = \alpha \ , \tag{B.1}$$

where α is the significance level of the limit, and $P(c|\mu)$ the PDF of the Poisson distribution with c observed and μ expected counts. Since the connection between k and s^U is monotonic,

the median expected exclusion limit can be obtained by considering, for a given expected background b, the median expected count number k. Since k can only acquire discrete values, this introduces jumps in the projected upper limits as function of b, which are clearly visible in Fig. 5.

The minimum number of events k that leads to a signal detection with a significance level of (at least) α, above an expected background of b events, is given by the smallest k that satisfies the inequality

$$\sum_{k' \geq k} P(k'|b) \leq \alpha \, , \tag{B.2}$$

which we call here k_{th}. The expected discovery limit of the signal for the given sensitivity level is now given by the smallest s that corresponds to a median count of k_{th}.

B.2 Maximum likelihood ratio method

For general likelihood functions with n free parameters, we can define the MLR test statistic (see Ref. [3] for details)

$$TS_{\mathcal{D}}(\theta_1) = -2 \ln \frac{\max_{\theta'_2,\ldots,\theta'_n \geq 0} \mathcal{L}(\mathcal{D}|\theta_1, \theta'_2, \ldots, \theta'_n)}{\max_{\theta'_1,\ldots,\theta'_n \geq 0} \mathcal{L}(\mathcal{D}|\theta'_1, \ldots, \theta'_n)} \, . \tag{B.3}$$

We fix here only one parameter, θ_1, and maximize w.r.t. the remaining ones. We define the modified test statistic

$$q_{\mathcal{D}}(\theta_1) = \begin{cases} TS_{\mathcal{D}}(\theta_1) & \text{if } \theta_1 > \hat{\theta}_1(\mathcal{D}) \\ 0 & \text{otherwise} \end{cases} \, , \tag{B.4}$$

where $\hat{\theta}_1 \geq 0$ is the maximum likelihood estimator given data $\mathcal{D}$. For a given data set, an one-sided confidence interval that corresponds to the desired upper limits is given by

$$C = \{\theta_1 \geq 0 \mid q_{\mathcal{D}}(\theta_1) \leq t(\theta_1)\} \, , \tag{B.5}$$

where the threshold $t(\theta_1)$ is in general a function of θ_1 and depends on the aspired significance level α. The threshold must be set such that C has correct coverage properties. This means that C should cover the true value of θ_1 in $1 - \alpha$ of the cases. In the large-sample limit, asymptotic formulae for the statistical distribution of $q_{\mathcal{D}}(\theta_1)$ are available [3]. However, in the small-sample regime, MC simulations are required to derive appropriate threshold values for $t(\theta_1)$. Note that C can be the empty set in some cases, which corresponds to downward fluctuations of the background. Although there are numerous ways to deal with this situations [23, 25], this is not problematic for the purposes of the present work. Note that, due to the discreteness of Poisson processes, $t(\theta_1)$ is in general not a continuous function of θ_1 or the other background parameters.

In the main text, we usually show median limits obtained from a large set of data realizations with $\theta_1 = 0$, using a $t(\theta_1)$ that is derived from MC simulations.

Expected discovery limits are derived in a similar way. We first find the threshold value t_{th} that corresponds to a test of the hypothesis $\theta_1 = 0$ with the significance level of α,

$$P(TS_{\mathcal{D}}(\theta_1 = 0) \geq t_{\text{th}}|\theta_1 = 0) \leq \alpha \, . \tag{B.6}$$

The discovery limit is then given by the smallest value of θ_1 that leads to a detection in at least 50% of the cases, namely we search for the θ_1^D that satisfies

$$P(TS_{\mathcal{D}}(\theta_1 = 0) \geq t_{\text{th}}|\theta_1 = \theta_1^D) \geq 0.5 \ . \tag{B.7}$$

Note that t_{th} and hence θ_1^D are not necessarily smooth functions of the background parameters $\theta_2, \ldots, \theta_n$, if the discreteness of the Poisson likelihood plays a role.

C Technical calculations

We present here some more details about derivations of equations related to discovery limits as well as the treatment of background systematics, used in the main part of the paper.

C.1 Expected discovery limits

Given $b \ll 1$ expected background events, one can derive an approximate discovery limit for the number of required signal events s and statistical significance α, by solving

$$P(s + b|b) = \frac{e^{-b}b^{s+b}}{\Gamma(s + b + 1)} = \alpha \ , \tag{C.1}$$

for s. Here, $P(s + b|b)$ is the continuum version of the Poisson probability mass function, with $s + b$ observed events while b are expected.

We compare this expression with Eq. (3.6) (we use the symbols s and b for simplicity). In the limit $b \ll s$ it can be written as

$$s \ln\left(\frac{s}{b}\right) - s = \frac{Z^2}{2} \ . \tag{C.2}$$

Now, one can consider the first two terms of the expansion of α in $1/Z$,

$$\ln\frac{1}{\alpha} \simeq \frac{Z^2}{2} + \frac{1}{2}\ln 2\pi Z^2 \ , \tag{C.3}$$

which can be substituted into the right-hand side of Eq. (C.2). The large-s approximation to the log of the gamma function, Stirling's formula, reads (we use $b \ll s$)

$$\ln\Gamma(s + 1) \approx s \ln s - s + \frac{1}{2}\ln 2\pi s \ , \tag{C.4}$$

which can be substituted in the left-hand side of Eq. (C.2). One can then rearrange the terms such that they read

$$\frac{b^s}{\Gamma(s + 1)} = \alpha \cdot \sqrt{\frac{Z^2}{s}} \ . \tag{C.5}$$

This has exactly the form shown in Eq. (3.8) (remember that $\lambda_i \simeq s_i$ in the low-background limit).

C.2 Examples with systematic errors

We will show here in some detail how to arrive at the results in Eqs. (4.3) and (4.8).

In the example leading to Eq. (4.3), we have a three component system with a signal, a background, and some third component which parametrizes variations in the background that are completely degenerate with the signal. The full Fisher matrix of the system is given by

$$
\mathcal{I} = \begin{pmatrix} \mathcal{I}_{11}^{\text{pois}} & \mathcal{I}_{12}^{\text{pois}} & \mathcal{I}_{11}^{\text{pois}} \\ \mathcal{I}_{12}^{\text{pois}} & \mathcal{I}_{22}^{\text{pois}} & \mathcal{I}_{12}^{\text{pois}} \\ \mathcal{I}_{11}^{\text{pois}} & \mathcal{I}_{12}^{\text{pois}} & \mathcal{I}_{11}^{\text{pois}} + \frac{1}{\xi_3^2} \end{pmatrix} , \tag{C.6}
$$

where we already used the various symmetry properties of the Fisher information matrix elements as well as the fact that $\Psi_1 = \Psi_3$. We are interested in the profiled 1×1 Fisher matrix where the parameter of interest is the signal parameter θ_1, and we have removed θ_2 and θ_3. This profiled Fisher matrix is given by

$$
\widetilde{\mathcal{I}} = \mathcal{I}_{11}^{\text{pois}} - \begin{pmatrix} \mathcal{I}_{12}^{\text{pois}} \\ \mathcal{I}_{11}^{\text{pois}} \end{pmatrix}^T \begin{pmatrix} \mathcal{I}_{22}^{\text{pois}} & \mathcal{I}_{12}^{\text{pois}} \\ \mathcal{I}_{12}^{\text{pois}} & \mathcal{I}_{11}^{\text{pois}} + \frac{1}{\xi_3^2} \end{pmatrix}^{-1} \begin{pmatrix} \mathcal{I}_{12}^{\text{pois}} \\ \mathcal{I}_{11}^{\text{pois}} \end{pmatrix} . \tag{C.7}
$$

It is straightforward to invert the 2×2 matrix analytically, and one can show that the inverse of the profiled Fisher matrix can be written in the simple form

$$
\frac{1}{\widetilde{\mathcal{I}}} = \xi_3^2 + \frac{1}{\mathcal{I}_{11}^{\text{pois}} - (\mathcal{I}_{22}^{\text{pois}})^{-1}(\mathcal{I}_{12}^{\text{pois}})^2} . \tag{C.8}
$$

If we now identify $\sigma_1^2(\boldsymbol{\theta}) = \widetilde{\mathcal{I}}^{-1}$, and $(\sigma_1^{\text{pois}})^2(\boldsymbol{\theta}) = (\mathcal{I}_{11}^{\text{pois}} - (\mathcal{I}_{22}^{\text{pois}})^{-1}(\mathcal{I}_{12}^{\text{pois}})^2)^{-1}$, we arrive at Eq. (4.3). Note that the latter is just the profiled Fisher information that we would have obtained in absence of the third component, or equivalently in the limit $\xi_3 \to \infty$.

The derivation of Eq. (4.8) follows a similar pattern, but is technically slightly more involved. Again, we are interested in the profiled 1×1 Fisher information for the signal component. It is here useful to associate the index $i = 0$ with the signal component, and the indices $i = 1, \dots, N$ with the discrete energies from Eq. (4.6). If we think about the underlying full Fisher matrix of the system in the block form shown in Eq. (2.8), then the components A (associated with $i = 0$), B and C are given by

$$
A = \sum_{i=1}^{N} \Delta E_i \frac{\Psi_1(E_i)^2}{\Phi(E_i)} , \tag{C.9}
$$

where we discretized the integral,

$$
C_i = \Delta E_i \frac{\Psi_1(E_i)\Psi_2(E_i)}{\Phi(E_i)} = \Delta E_i \Psi_1(E_i) , \tag{C.10}
$$

where $i = 1, \dots, N$ and we used in the second step that $\Phi(E) = \Psi_2(E)$, and

$$
B = \delta_{ij}\Delta E_i \Phi(E_i) + \Sigma_\delta^{-1} , \tag{C.11}
$$

where we included the inverse of the covariance matrix for ξ_i. Then, the profiled Fisher information can be written as

$$
\begin{aligned}
\widetilde{\mathcal{I}}_{11} &= \sum_{i=1}^{N} \Delta E_i \frac{\Psi_1(E_i)^2}{\Phi(E_i)} - \sum_{i,j=1}^{N} \Delta E_i \Delta E_j \Psi_1(E_i)\Psi_1(E_j) \left[\mathrm{diag}(\Delta E_i \Psi_2(E_i)) + \boldsymbol{\Sigma}_\delta^{-1}\right]_{ij}^{-1} \\
&= \sum_{i=1}^{N} \Delta E_i \frac{\Psi_1(E_i)^2}{\Phi(E_i)} - \sum_{i,j=1}^{N} \sqrt{\frac{\Delta E_i \Delta E_j}{\Phi(E_i)\Phi(E_j)}} \Psi_1(E_i)\Psi_1(E_j) \left[\frac{\boldsymbol{\Sigma}_\delta'}{1 + \boldsymbol{\Sigma}_\delta'}\right]_{ij} \\
&= \sum_{i,j=1}^{N} \sqrt{\frac{\Delta E_i \Delta E_j}{\Phi(E_i)\Phi(E_j)}} \Psi_1(E_i)\Psi_1(E_j) \left[\frac{1}{1 + \boldsymbol{\Sigma}_\delta'}\right]_{ij} .
\end{aligned}
\tag{C.12}
$$

Here, we used the definitions $(\boldsymbol{\Sigma}_\delta)_{ij} = \Sigma_\delta(E_i, E_j)$ and $(\boldsymbol{\Sigma}_\delta')_{ij} = \sqrt{\Delta E_i \Delta E_j \Phi(E_i)\Phi(E_j)}\, \Sigma_\delta(E_i, E_j)$. In the first step, we rearrange some factors of ΔE_i and $\Phi_1(E_i)$, and use the general matrix relation $(1 + M^{-1})^{-1} = M/(1 + M)$. In the second step, we include the matrix identity in the form $\delta_{ij} = [(1 + \boldsymbol{\Sigma}_\delta')/(1 + \boldsymbol{\Sigma}_\delta')]_{ij}$ in the first summation which helps to further collapse the whole expression. The last line is after some more rewriting equivalent to Eq. (4.8).