



PUBLISHED FOR SISSA BY SPRINGER

RECEIVED: July 30, 2014  
REVISED: September 29, 2014  
ACCEPTED: October 20, 2014  
PUBLISHED: November 6, 2014

# Improving estimates of the number of ‘fake’ leptons and other mis-reconstructed objects in hadron collider events: BoB’s your UNCLE<sup>1</sup>

Thomas P.S. Gillam and Christopher G. Lester

*Dept of Physics, Cavendish Laboratory,  
JJ Thomson Avenue, Cambridge, CB3 0HE, U.K.*

*E-mail:* [gillam@hep.phy.cam.ac.uk](mailto:gillam@hep.phy.cam.ac.uk), [lester@hep.phy.cam.ac.uk](mailto:lester@hep.phy.cam.ac.uk)

**ABSTRACT:** We consider current and alternative approaches to setting limits on new physics signals having backgrounds from misidentified objects; for example jets misidentified as leptons,  $b$ -jets or photons. Many ATLAS and CMS analyses have used a heuristic “matrix method” for estimating the background contribution from such sources. We demonstrate that the matrix method suffers from statistical shortcomings that can adversely affect its ability to set robust limits. A rigorous alternative method is discussed, and is seen to produce fake rate estimates and limits with better qualities, but is found to be too costly to use. Having investigated the nature of the approximations used to derive the matrix method, we propose a third strategy that is seen to marry the speed of the matrix method to the performance and physicality of the more rigorous approach.

**KEYWORDS:** Lepton production, Hadron-Hadron Scattering, Jets, Jet physics

**ARXIV EPRINT:** [1407.5624](https://arxiv.org/abs/1407.5624)

<sup>1</sup>‘BoB Method’ is short for ‘Best bits of Both existing fake rate estimation **M**ethods’. UNCLE stands for “**U**n-biased **C**onfidence **L**imit **E**valuator”. Both are somewhat artificial acronyms designed to make the name of Method C in table 1 memorable.

---

**Contents**

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Overview of fake estimation procedures and limit setting</b>	<b>3</b>
2.1	Method A: the “matrix method”	3
2.1.1	Events with only one lepton	4
2.1.2	Events with multiple leptons	5
2.1.3	Limit setting	6
2.2	Method B: an extended likelihood method	7
2.2.1	Choice of parameterisation	7
2.3	Method C: maximum likelihood estimate	9
<b>3</b>	<b>Comparisons using frequentist limits</b>	<b>9</b>
3.1	Simple scenario — two leptons, two categories	10
3.2	Harder scenario — two leptons, eight categories	10
<b>4</b>	<b>Conclusions</b>	<b>12</b>
<b>A</b>	<b>Origin of matrix method approximation</b>	<b>14</b>
A.1	Single lepton and single category	14
A.2	Multiple leptons and multiple categories	15
<b>B</b>	<b>Tables of parameters</b>	<b>16</b>

---

**1 Introduction**

Many precision measurements and searches for new physics employ signal regions which have a significant source of background coming from ‘fake’ objects. A typical concrete example is that of leptons, which can be faked by a mis-reconstructed jet. Alternatively one can consider jets faking  $b$ -jets, for which the matrix method was used in [1], or even faking photons. In this article the term ‘lepton’ shall be used throughout, but all statements made are general to other types of object. Properties whose distributions differ for ‘fake’ and ‘real’ objects have been used to underpin data-driven methods of fake rate estimation, one of the most prevalent of which during the first data-taking run of the LHC has been the so-called ‘matrix method’ [2] used by ATLAS in [3–31], and by CMS in [32–34] based, apparently, on a description in [35].

The matrix method is the first of three ways of determining fake rates that are compared in this paper. We shall sometimes refer to it for short Method A to facilitate easy comparison with the later Methods B and C. (For quick reference see table 1.)

Method	Cost of computation	Limit quality	Other names
<b>A</b>	low	poor	‘Matrix Method’ [2]
<b>B</b>	very high	very good	‘Likelihood Matrix Method’
<b>C</b>	quite low	good	‘BoB Method’

**Table 1.** An overview of the three methods discussed in this paper, and their relative strengths and weaknesses. ‘Limit quality’ refers to whether  $CL_{s+b}$  limits tend to have the correct frequentist coverage properties, and also avoid unnecessary over-coverage.

Method A makes use of the fact that fake and real leptons tend to differ in their degree of ‘isolation’.<sup>1</sup> Using cuts on isolation (and to a lesser extent other variables) leptons are categorised as either ‘tight’ or ‘loose’, the former being largely synonymous with ‘more isolated’ and the latter with ‘less isolated’. In this paper it is shown that the way in which the resulting matrix method derived background estimates are typically used in existing SUSY searches can give rise to confidence limits that are unstable (highly variable) indicating they are making non-optimal use of the data. More specifically, over the course of many independent experiments one expects to find a distribution of limits from Method A which has a larger variance (is more widely spread out) than the distribution of limits coming from Methods B and C discussed later. In addition, Method A can produce unphysically negative estimates for fake rates that should be bounded below by zero.

The weaknesses of the matrix method stem from the presence of heuristic, non-mathematical steps in its derivation. Heuristic methods are often useful, but can be problematic if their underlying assumptions are sometimes invalid; we show this is the case with the matrix method. Methods lacking these deficiencies are in principle trivial to construct. In Method B we describe an example of such a method in which a single likelihood is used to perform both the background estimation and the limit setting. Whilst this can be considered the optimal approach, it is shown to be computationally expensive in cases where objects are divided into many categories.<sup>2</sup> In order to marry the best of Methods A and B, we then propose Method C. It is intended that Method C be usable as a drop-in replacement for Method A in the contexts in which the latter has previously been used by ATLAS or CMS. Method C, like Method A, it is partly heuristic (for speed) and so is justified pragmatically. However, careful choice of the approximations it contains allows it to always give physical limits whose distributions very closely resemble the optimal (but prohibitively expensive) Method B.

Note that fake rate estimates in LHC analyses are likely to find themselves being used as part of a  $CL_s$  frequentist limit setting procedures, since these are endemic in ATLAS and CMS papers. Such usage requires a likelihood for a set of parameters given observed data; in the case where one counts events in just one region, it typically takes the form of

<sup>1</sup>The ‘isolation’ of a charged object is, in this case, a measure of the amount of activity found close to the track left by the object in the inner detector.

<sup>2</sup>Many categories would be required if reconstruction efficiencies varied as a function of, say, detector rapidity or lepton transverse momentum.

a Poisson distribution having some mean. Background estimates in an analysis are then interpreted as an auxiliary measurement which constrains this mean through additional terms in the likelihood. In this context, the fake rate ought to be an estimate of the expected number of events from the fake background process in the signal region, given data collected *outside* of the signal region. This is not strictly adhered to in the matrix methods, and so is one of the general places in which Methods B and C improve on A.

## 2 Overview of fake estimation procedures and limit setting

Events collected into some signal region are defined in terms of the numbers of leptons they contain. A cut on some measure of quality, for example isolation, distinguishes a given lepton as *loose* ( $l$ ) or *tight* ( $t$ ), where  $l \cup t \equiv \tilde{l}$  and  $l \cap t = \emptyset$ . Each lepton will also be regarded as either *real* ( $r$ ) or *fake* ( $f$ ), depending on whether it is a correctly reconstructed lepton, or for example a mis-reconstructed heavy flavour jet. According to the precise selection, a certain number of tight and loose leptons will be required for an object to make it into a signal region — those that do are described as *tight events* ( $T$ ), and those that do not but *could* be made to pass the selection for some permutation of  $t$  and  $l$  on its constituent leptons are denoted as loose events ( $L$ ), where as before  $L \cup T \equiv \tilde{L}$  and  $L \cap T = \emptyset$ .

A core concept in all methods considered here is that of the real and fake *efficiencies*, respectively defined to be  $\varepsilon_r \equiv P(t|r\tilde{l})$  and  $\varepsilon_f \equiv P(t|f\tilde{l})$ . For convenience we will also use  $\bar{\varepsilon}_r \equiv 1 - \varepsilon_r = P(l|r\tilde{l})$  and  $\bar{\varepsilon}_f \equiv 1 - \varepsilon_f = P(l|f\tilde{l})$ . Typically these quantities are measured in additional control regions, and could be subdivided according to kinematic quantities, such as lepton  $p_T$ . In this text such categories will be labelled  $\omega_1, \omega_2, \dots$ , with the efficiencies gaining an additional subscript e.g.  $\varepsilon_{\omega_1 r}$ .

For a given event containing  $m$  leptons, each lepton is observed to be either  $l$  or  $t$ , and will have some category  $\omega_i$ . If there are  $N_\omega$  possible categories for each lepton, then the number of measurable *event categories* will be  $N_\Omega = (2 \times N_\omega)^m$ .<sup>3</sup> Each of these will correspond to an event that is either  $L$  or  $T$ .

Experimentally, one counts how many events fall into each of the  $N_\Omega$  sub-regions, yielding the set of integers  $\{n_{\Omega_i}\}$ . For the purpose of the physics analysis being performed, one might be interested in the total number of tight events,  $n_T = \sum_{\Omega_i \subset T} n_{\Omega_i}$ . Usually this is the quantity with which a limit on the cross section of a new physics model is placed.

The observed numbers of events are often assumed to be the particular values of a Poisson distributed random variable. That is, one can have  $n_T \sim \text{Poiss}(\nu_T)$ ; in general the indices on the rate  $\nu$  correspond to those on the observation  $n$ .

### 2.1 Method A: the “matrix method”

This section attempts to document Method A, the matrix method, in more detail than has previously been done, and in its most general form. As mentioned previously it is a somewhat heuristic method, but its assumptions shall be interpreted on a firmer statistical footing in a subsequent section.

---

<sup>3</sup>If the number of leptons can differ between events, one introduces an appropriate sum over  $m$ .

### 2.1.1 Events with only one lepton

Consider first a simplified scenario where each event has exactly one lepton;  $n_T$  tight and  $n_L$  loose events are observed. The key relation is then that

$$\begin{pmatrix} \langle n_T \rangle \\ \langle n_L \rangle \end{pmatrix} = \begin{pmatrix} \varepsilon_r & \varepsilon_f \\ \bar{\varepsilon}_r & \bar{\varepsilon}_f \end{pmatrix} \begin{pmatrix} n_R \\ n_F \end{pmatrix}, \quad (2.1)$$

where  $n_R$  and  $n_F$  are the number of the observed events which are real and fake, respectively. In this context,  $\langle n_L \rangle = \mathbb{E}[n_L|n_R, n_F]$ , and similarly for  $\langle n_T \rangle$ . The result follows by considering the real/fake event counts to be random variables following a Poisson distribution, which are then further divided into tight and loose components according to a binomial distribution using the probabilities contained in the efficiencies. It can be noted that equation (2.1) is similar to a relation between the means of Poisson distributions

$$\begin{pmatrix} \nu_T \\ \nu_L \end{pmatrix} = \begin{pmatrix} \varepsilon_r & \varepsilon_f \\ \bar{\varepsilon}_r & \bar{\varepsilon}_f \end{pmatrix} \begin{pmatrix} \nu_R \\ \nu_F \end{pmatrix}. \quad (2.2)$$

This is used later when discussing Method B and Method C, but for now we shall proceed with equation (2.1). This equation may legitimately be inverted

$$\begin{pmatrix} n_R \\ n_F \end{pmatrix} = \frac{1}{\varepsilon_r - \varepsilon_f} \begin{pmatrix} \bar{\varepsilon}_f & -\varepsilon_f \\ -\bar{\varepsilon}_r & \varepsilon_r \end{pmatrix} \begin{pmatrix} \langle n_T \rangle \\ \langle n_L \rangle \end{pmatrix}, \quad (2.3)$$

provided that  $\varepsilon_r \neq \varepsilon_f$ . Given the model assumptions that were made, the steps described hitherto all hold water on mathematical grounds. In contrast, the next step that is usually taken to motivate Method A is quite arbitrary, and is justified largely on grounds that it is effective in situations with large numbers of events, rather than because it is meaningful in general.<sup>4</sup> This ‘heuristic’ step consists of the removal of the expectation brackets from the right hand side of equation (2.3) and the re-interpretation of the terms on its left hand side as a pair of quantities  $\hat{n}_R$  and  $\hat{n}_F$  as follows:

$$\begin{pmatrix} \hat{n}_R \\ \hat{n}_F \end{pmatrix} = \frac{1}{\varepsilon_r - \varepsilon_f} \begin{pmatrix} \bar{\varepsilon}_f & -\varepsilon_f \\ -\bar{\varepsilon}_r & \varepsilon_r \end{pmatrix} \begin{pmatrix} n_T \\ n_L \end{pmatrix}. \quad (2.4)$$

What are  $\hat{n}_R$  and  $\hat{n}_F$ ? They depend on  $n_T$  and  $n_L$  and so are functions of the data, and may be regarded as estimators — but estimators for what? It is shown in appendix A that under some additional assumptions, and for certain values of  $n_T$  and  $n_L$ , they turn out to be maximum likelihood estimators for  $n_R$  and  $n_F$  given knowledge of  $n_T$  and  $n_L$  (i.e. estimators for  $\langle n_R \rangle \equiv \mathbb{E}[n_R|n_T, n_L]$  and  $\langle n_F \rangle \equiv \mathbb{E}[n_F|n_T, n_L]$ ).<sup>5</sup> Nonetheless, and in the absence of anything better, Method A instead uses  $\hat{n}_R$  and  $\hat{n}_F$  as estimators for the unknown and unknowable *actual* rates of real and fake events,  $\nu_R$  and  $\nu_F$ .

<sup>4</sup>We shall see in fact that it is *not* meaningful when the numbers in question are small.

<sup>5</sup>Note, as shown, that these expectation values are necessarily conditioned on different things to the expectations values  $\langle n_T \rangle$  and  $\langle n_L \rangle$  seen earlier.

Note that these estimators are sometimes pretty bad as equation (2.4) allows terms on its left hand side to become unphysically negative.<sup>6</sup> This happens in real analyses (e.g. [5]) creating problems that need to be solved by *ad-hoc* methods. Both Method B and Method C have the benefit of avoiding such problems.

Finally, Method A obtains its desired goal, the definition of an estimator for the expected number of fake events in the signal region,  $\hat{n}_{TF}$ , motivated by equation (2.1) with the replacement  $n_F \rightarrow \hat{n}_F$ , where  $\hat{n}_F$  is the estimator obtained above in (2.4). This results in:

$$\hat{n}_{TF} = \frac{\varepsilon_f}{\varepsilon_r - \varepsilon_f} (\varepsilon_r(n_T + n_L) - n_T). \quad (2.5)$$

Again, note the problems with this method that, even if  $\varepsilon_r > \varepsilon_f$  (as must be the case for a useful definition of  $t$  and  $l$ ), equation (2.5) can yield  $\hat{n}_{TF} < 0$ , an unphysical result which is symptomatic of the earlier “sleight of hand”.

This concludes our description of how Method A is used in single lepton events to calculate a number which is used as if it were an estimate of the expected rate of fakes in the signal region. We will now describe how the same method is extended for use in events with more than one lepton, which has previously not been documented in detail.

### 2.1.2 Events with multiple leptons

Consider an event with two leptons, where each lepton can be in one of a number of categories  $\{\omega_i\}$ . One may define quantities such as  $n_{tl}$ , the number of events with the first<sup>7</sup> lepton tight and the second loose — others are defined similarly. In order to include the possible categories for each lepton, event counts such as  $n_{tt}$  must be further subdivided to take into account all combinations:

$$n_{tt} = n_{tt}^{\omega_1\omega_1} + n_{tt}^{\omega_1\omega_2} + \dots \quad (2.6)$$

$$= \sum_{i,j} n_{tt}^{\omega_i\omega_j}. \quad (2.7)$$

In this notation,  $n_{tt}^{\omega_1\omega_2}$  indicates the number of events with two tight leptons, where the first is in category  $\omega_1$ , and the second in  $\omega_2$ .

The analogous relation to equation (2.1) is then

$$\left\langle n_{\mathcal{T}\beta_1\beta_2}^{\omega_i\omega_j} \right\rangle = \phi_{\omega_i\beta_1}^{\alpha_1} \phi_{\omega_j\beta_2}^{\alpha_2} n_{\mathcal{R}\alpha_1\alpha_2}^{\omega_i\omega_j}, \quad \text{where} \quad (2.8)$$

$$\phi_{\omega_i\{t,l\}}^{\{r,f\}} = P(\{t,l\}|\{r,f\}\omega_i\tilde{l}), \quad (2.9)$$

$$\phi_{\omega_i} = \begin{pmatrix} \varepsilon_{\omega_i r} & \varepsilon_{\omega_i f} \\ \bar{\varepsilon}_{\omega_i r} & \bar{\varepsilon}_{\omega_i f} \end{pmatrix}, \quad (2.10)$$

where summation over repeated upper and lower indices is implied where appropriate.<sup>8</sup> The identifier  $n$  has been replaced for clarity with the symbols  $n_{\mathcal{T}}$  and  $n_{\mathcal{R}}$ , depending on

<sup>6</sup>For example, consider the case where  $n_T = 7, n_L = 1, \varepsilon_r = \frac{4}{5}$ , and  $\varepsilon_f = \frac{1}{5}$ . One can then show that  $\hat{n}_R = 9$ , and  $\hat{n}_F = -1$ .

<sup>7</sup>The definition of “first” can depend on the analysis. Often it is chosen to be the hardest according to  $p_T$ .

<sup>8</sup>In this case a sum should *not* be carried out over  $\omega_i$  or  $\omega_j$ , since they appear on the left-hand side. Despite some notational and behavioural similarities, these objects are not tensors!

whether the accompanying indices pertain to tight/loose-ness, e.g.  $n_{\mathcal{T}tl} = n_{tl}$ , or real/fake-ness, so  $n_{\mathcal{R}rf} = n_{rf}$ . Each Greek lower index of  $n_{\mathcal{T}}$  hence takes values in  $\{t, l\}$ , while each Greek lower index of  $n_{\mathcal{R}}$  takes values in  $\{r, f\}$ . The index on these indices corresponds to the which lepton is being described; i.e. in equation (2.10), the value of  $\alpha_2$  represents whether the second lepton is either real or fake.

The matrix representation for  $\phi_{\omega_i}$  shown in the last line of equation (2.10) is not needed to understand this equation, but is required when considering the background estimate for events that are both tight and fake (it is what still identifies this as the “matrix method”, despite the new notation).

The estimate for the expected number of *events* that are fake is then  $\hat{n}_{\mathcal{T}\alpha_1\alpha_2}^F$ , where

$$\hat{n}_{\mathcal{T}\nu_1\nu_2}^F = \sum_{i,j} \left( \phi_{\omega_i\nu_1}^{\mu_1} \phi_{\omega_j\nu_2}^{\mu_2} \zeta_{\mu_1\mu_2}^{\beta_1\beta_2} \phi_{\omega_i\beta_1}^{-1\alpha_1} \phi_{\omega_j\beta_2}^{-1\alpha_2} n_{\mathcal{T}\alpha_1\alpha_2}^{\omega_i\omega_j} \right), \quad (2.11)$$

where the  $\zeta$  object is responsible for defining what is meant by a fake event. For example, if  $rr \equiv R$  and  $\{rf, fr, ff\} \equiv F$  then one would choose  $\zeta_{12}^{12} = \zeta_{21}^{21} = \zeta_{22}^{22} = 1$ , and all other components 0. There is in fact a redundancy in the indices, in that all non-zero components have the  $i^{\text{th}}$  lower index the same as the  $i^{\text{th}}$  upper index. In general therefore, for the case with any number of leptons

$$\zeta_{\mu_1\mu_2\cdots}^{\beta_1\beta_2\cdots} = \delta_{\mu_1}^{\beta_1} \delta_{\mu_2}^{\beta_2} \cdots h(\beta_1, \beta_2, \dots), \quad (2.12)$$

where  $\delta_i^j$  is the Kronecker delta, and  $h(\beta_1, \dots)$  is a function of the indices that is 1 for a fake combination, and 0 for a real combination.

In order to estimate the number of events contained within  $\hat{\mathcal{T}}^F$ , from equation (2.11), that are tight, one sums the appropriate component(s). For example, a simple analysis selecting final states with exactly two leptons might define  $T \equiv tt$  (i.e. the number of tight *events* would now be denoted  $n_T \equiv n_{tt}$ ), and all other possibilities to be  $L$ . In this case the  $\hat{\mathcal{T}}_{tt}^F$  component is the estimate of the number of events that are both tight and fake. For the completely general case with events containing arbitrary numbers of leptons, additional terms and indices are added as necessary to the equations in this section.

### 2.1.3 Limit setting

ATLAS and CMS analyses use the  $CL_s$  method [36] to place an upper limit on the event rate (in the sense of the mean of a Poisson distribution, which controls the appearance of events in a signal region) of new physics processes.

In the context of limit setting, the output from the matrix method is treated on a par with those irreducible background components estimated from Monte Carlo simulated (MC) samples. Once the central value is estimated as described in section 2.1, uncertainties in the measured efficiencies, as well as a statistical uncertainty, can be propagated in the usual way by taking a derivative [37]. The background mean  $\bar{b}$  and uncertainty  $\sigma_b$  are fed into a joint likelihood for the signal and background rates,  $\mu$  and  $b$ , given the number of events observed in the signal region  $n_T$ . In the case with only one background source it takes the form

$$\mathcal{L}(\mu, b|n_T) = \text{Poiss}(n_T; \mu + b) \text{Gauss}(\bar{b}; b, \sigma_b). \quad (2.13)$$

When setting the limit, the nuisance parameter  $b$  is profiled away in the usual way to form the test statistic  $q_\mu$ ,

$$q_\mu = \begin{cases} -2 \ln \left( \frac{\mathcal{L}(\mu, \hat{b} | n_T)}{\mathcal{L}(\hat{\mu}, \hat{b} | n_T)} \right) & \mu > \hat{\mu} \\ 0 & \mu \leq \hat{\mu} \end{cases} . \quad (2.14)$$

Confidence intervals ( $CL_s$  or  $CL_{s+b}$ ) at some level can then be formed by following the recipe outlined in [36].

## 2.2 Method B: an extended likelihood method

Whilst Method A can suffer from under-coverage, as subsequently discussed in section 3, this can largely be avoided for a purely data-driven background if the *full* likelihood, including all data used to make the measurement, is used in the limit setting procedure. That is, one should use

$$\mathcal{L}(\mu, \boldsymbol{\theta} | n_t, n_l, n_{tt}, \dots), \quad (2.15)$$

where  $\boldsymbol{\theta}$  represents the set of nuisance parameters. If the leptons can fall into one of several categories, quantities should be replaced with the separate terms from equation (2.7). Each of these quantities can be considered as an independent random variable with a Poisson distribution. The means of these Poisson distributions will be denoted as functions of the parameters e.g.  $\nu_{tt}^{\omega_1 \omega_1}(\mu, \boldsymbol{\theta})$ ; the likelihood then factorises and takes a form similar to equation (2.13)

$$\mathcal{L}(\mu, \boldsymbol{\theta} | \dots, n_{tt}^{\omega_1 \omega_1}, \dots) = \text{Pois}(n_{tt}^{\omega_1 \omega_1}; \nu_{tt}^{\omega_1 \omega_1}(\mu, \boldsymbol{\theta})) \cdots P(\bar{\boldsymbol{\theta}} | \boldsymbol{\theta}). \quad (2.16)$$

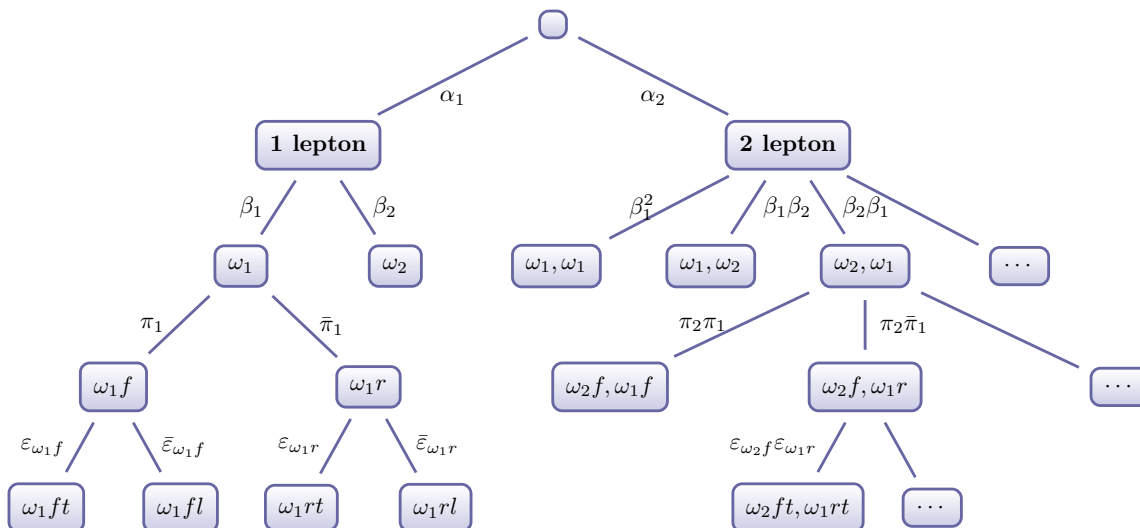
The final term represents constraints placed on the nuisance parameters by external measurements.

### 2.2.1 Choice of parameterisation

The efficacy of any likelihood method depends on a sensible choice of parameterisation. The parameterisation must completely describe how events from both signal and background are expected to be divided between the different event categories without over-parameterising. For example, one *could* directly use  $\nu_{tt}^{\omega_1 \omega_1}$  etc. as the free parameters  $\boldsymbol{\theta}$ , but this would remove all predictive power!

Other researchers [38] have investigated the possibility of applying a method that uses a similar parameterisation to the matrix method. This parameterisation uses the efficiencies described before, in addition to the rates separated both by object category and real/fake-ness. Whilst this has an advantage of making minimal assumptions about how a given background process distributes itself between these categories, it does lead to a very large parameter space. For example, even with three objects coming from only three possible categories, there are already 80 such parameters (before considering efficiencies). Since any form of prediction will require a maximisation of the likelihood over this input parameter





**Figure 1.** The probability tree in this figure illustrates the model used to parameterise fake and real lepton production, as used in Method B and Method C. The left-most branch is complete, the others are not as indicated by the presence of “...”. In general one could allow both for more lepton categories, as well as more leptons in the event. Note that  $\sum_{m=0}^{m_{\max}} \alpha_m = \sum_{i=1}^{N_\omega} \beta_i = 1$ , where  $m_{\max}$  is the largest number of leptons that can be produced in a given event. Additionally, the abbreviation  $\bar{\pi}_i = 1 - \pi_i$  is used.

space, and since such global maximisations become computationally more expensive as dimensionality increases, the authors have chosen to use an alternative parameterisation.

Diagrammatically, the parameterisation used in this work is displayed in figure 1. For every event that is generated, it is first decided how many leptons that event ought to contain. This is controlled by a set of parameters  $\{\alpha_m\}$ , each of which corresponds to the probability of forming an event with  $m$  leptons. As noted in the caption, these must sum to 1. For each lepton, a category  $\omega_i$  is assigned to it with probability  $\beta_i$ , and it is then further assigned to be either  $r$  with probability  $\pi_i$  or  $f$  with probability  $1 - \pi_i$ . Formally,  $\beta_i \equiv P(\omega_i|\tilde{l})$  and  $\pi_i \equiv P(f|\omega_i\tilde{l})$ . Efficiencies are then used in the usual way to select objects as being  $t$  or  $l$ .

Using these terms, together with one extra non-negative parameter denoting the mean of the Poisson distribution controlling the total production of tight *events*,<sup>9</sup> one can compute the terms such as  $\nu_{tt}^{\omega_1\omega_1}$  in equation (2.16). It should be noted that one of these trees must exist for every separate ‘component’ that is being fitted — that is, at least one for the hypothesised signal process and one for the fake component of the background, and then optionally one or more for other background components that have been estimated using MC samples.

<sup>9</sup>One could alternatively use the overall production of  $\tilde{L}$  events, however it is essential to have the rate of  $T$  events as a parameter for any signal component, since this is the quantity upon which one wishes to place a limit.

### 2.3 Method C: maximum likelihood estimate

It is later found that Method B is computationally intractable for more than very simple systems. As such, we propose a third method that is found to keep many of the desirable properties of Method B, but also with a much reduced computation time more similar to Method A. This is achieved by using a simple likelihood for limit setting, as in Method A, but feeding it with the true maximum likelihood estimate (MLE) fake rate.

To form an upper limit with Method C, one does the following. Firstly, for the observed data, maximise the likelihood expressed in equation (2.16) for all nuisance parameters. As mentioned in the discussion for Method B, this likelihood should contain sufficient parameters to describe the signal process, fake background, and any other real backgrounds. The output from this that shall be used is the MLE fake rate with an estimated uncertainty. This uncertainty represents both an uncertainty with which the efficiencies are known, as well as statistical limitations of the observed data. It is estimated with the MINOS method [39], by taking the values of the fake rate where the minimum of the negative log likelihood with respect to the *remaining* parameters increases by 0.5 from its minimum value. A limit is then placed using an expression identical to that in equation (2.13), where  $\bar{b}$  and  $\sigma_b$  take the aforementioned MLE fake rate and uncertainty.

## 3 Comparisons using frequentist limits

Using a toy event generator, written by the authors, datasets are produced using the same method as that which is depicted in figure 1, containing a mixture of ‘fake’ and ‘signal’ events. For each of several configurations, 19000 independent datasets were formed using the generator. Each of these was subsequently processed using Methods A, B and C. In all cases the necessary minimisation of a negative log likelihood was performed using the Minuit2 library [39]. The result are 95%  $CL_{s+b}$  and  $CL_s$  upper limits on the signal strength parameter.<sup>10</sup>

There is some discussion in the literature regarding how the incorporation of background components comprising a mean with some uncertainty affects the frequentist coverage properties of  $p$ -value limits [41]. In particular, when one is considering a background that is constrained e.g. from an MC sample, the acceptance region for the hypothesis test in the full Neyman construction will vary according to the value assumed by the nuisance parameter(s) controlling the strength of the background. In an approximated scheme, such as the profiling method used in the computation of  $CL_{s+b}$  and  $CL_s$ , the coverage can hence deviate from that nominally expected; potentially significantly if the background overestimates the data. Since both Methods A and C feed information into the likelihood in a similar way (and have the shortcoming that the likelihood used in the limit-setting procedure is not the likelihood for *all* the data), we should not be surprised if one or both methods under- or over-cover. It is hoped, however, that by virtue of the MLE fake rate being more ‘sensible’ than that from the matrix method, any deviations in coverage from that

<sup>10</sup>The  $p$ -values used to compute  $CL_s$  and  $CL_{s+b}$  are computed by performing pseudo-experiments, rather than using asymptotic methods [40], since it is known that the latter are only a good approximation for scenarios with a large number of events. In this work we focus on regions with low numbers of events.

nominally expected would be less extreme in Method C than with Method A. Method B should have the most accurate coverage, although it still might not be exactly correct due to the use of profiling. These expectations will be confirmed in the results which follow.

### 3.1 Simple scenario — two leptons, two categories

Firstly, a configuration is used that produces events always with exactly two leptons, each of which can be in one of two categories. There are separate configurations for a signal process, which produces only real leptons ( $\pi_1 = \pi_2 = 0$ ), and a fake process which produces only fake leptons ( $\pi_1 = \pi_2 = 1$ ). The full set of parameters can be found in appendix B in table 2. In each dataset, 100 events are produced using the tree in figure 1. As such the number of  $T$  events is approximately the sum of two Poisson random variables; one representing the signal component with mean 0.706, and another representing the fake background with a mean of 1.94.

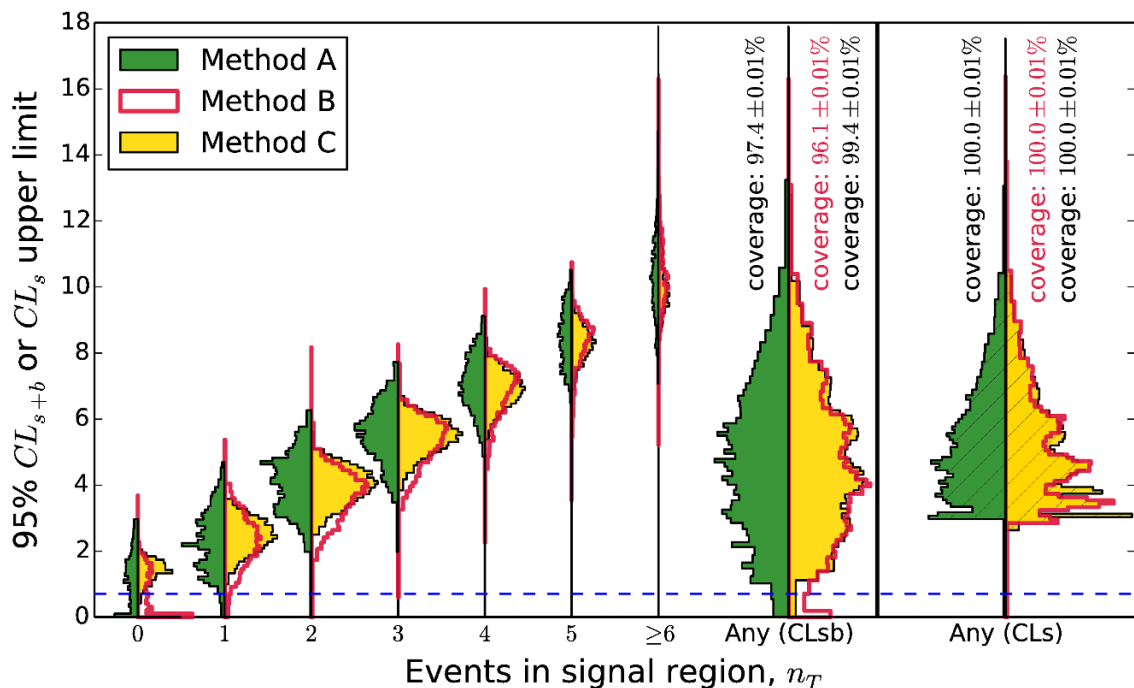
The  $CL_{s+b}$  and  $CL_s$  limits from each of the 19000 generated datasets is shown in figure 2. The  $CL_{s+b}$  limit is shown to have approximately correct coverage for Method A and Method B, but Method C over-covers; deviations from 95% at this level can only be justified on the grounds of the use of a profiled test statistic, rather than the full Neyman construction. This is further considered for the next example. There is also significant over-coverage in the  $CL_s$  limit, however this is expected due to the definition  $CL_s = \frac{CL_{s+b}}{1-CL_b}$ . In low statistics regimes, often  $(1 - CL_b) < 1$ , meaning that  $CL_s > CL_{s+b}$  by a potentially significant margin.

Furthermore, a division of the  $CL_{s+b}$  limit according to the number of events observed in the signal region,  $n_T$ , is also shown in figure 2. From this figure, whilst it can be seen that overall very similar limits are being placed by all three methods, in fact Method B tends to be most constraining due to its distributions showing longer lower tails. Method B and Method C are together significantly more constraining than Method A (signified by shorter upper tails), and are quite similar to each other; this is encouraging in justifying the use of Method C as an approximation to Method B. Method B is slightly more likely to place a tighter limit, as is to be expected since it makes optimal use of all available information.

A further comparison that can be made is of the fake rate that is the output of the matrix method in Method A, against the MLE of the fake rate obtained in Method B and Method C; this is shown in figure 3. The spread in the plot demonstrates the property that Method A can predict a negative fake rate, as seen in a significant portion of the generated datasets. It also shows that Method B produces fake rates that cluster more closely around the true value, even at low  $n_T$ .

### 3.2 Harder scenario — two leptons, eight categories

The simple scenario above has been extended to use eight categories instead of two. As per the parameterisation in figure 1, this involves the addition of 24 extra parameters — twelve each for the signal and fake background from the addition of six  $\beta$  and six  $\pi$  terms. The full set of parameters can be found in appendix B in table 3. As before, 100 events were generated in each dataset, corresponding to a signal rate of 0.748 and a fake background rate of 2.77.

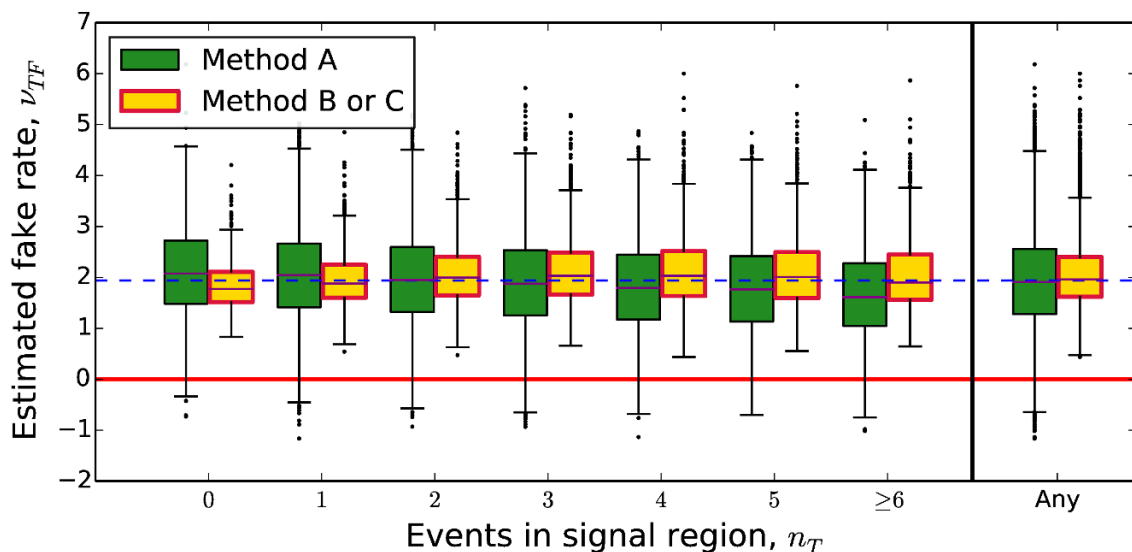


**Figure 2.** In the *simple scenario* described in the text (two leptons in two categories) computations using Method B are tractable. This figure therefore compares Methods A, B and C in that configuration. The 95%  $CL_{s+b}$  and  $CL_s$  upper limits on the rate of  $T$  signal events for each of 19000 independent toy datasets are histogrammed. For each ‘column’, histograms are made for each of Methods A, B and C and plotted back-to-back; that is A is plotted opposite to B and C, which overlap. The  $CL_{s+b}$  results are further divided into bins of observed  $n_T$ ; in all cases the area of each histogram is proportional to the number of toy datasets used to create it. The dashed blue line indicates the true signal production rate,  $\nu_{TR} = 0.706$ . The coverage of the observed limits of this truth rate are noted for the overall  $CL_{s+b}$  and  $CL_s$  results.

It was found that the increase in parameter space dimensionality was sufficient to increase the computation time for the likelihood maximisation to such an extent that producing limits with Method B became infeasible using the resources at the authors’ disposal; as such only results from Method A and Method C could be computed.

Figure 5 shows that the MLE fake rate for Method C is much more tightly constrained around the true value than the Method A estimate; moreover Method A gives even more significant deviations into negative values than with the simple scenario. Furthermore, as  $n_T$  increases, the median fake rate from Method A decreases slightly, whereas that from Method C is stable for low event counts, only increasing slightly for larger  $n_T$ ; the Method C behaviour seems more desirable here. Secondly, figure 4 shows that the  $CL_{s+b}$  limits derived in Method A suffer from under-coverage;<sup>11</sup> the upper limit only bounds the true rate 92% of the time rather than the expected 95%. Finally the upper tails of the

<sup>11</sup>At least within particle physics, under-coverage is typically considered a worse crime than over-coverage, as the related limits are then not conservative.

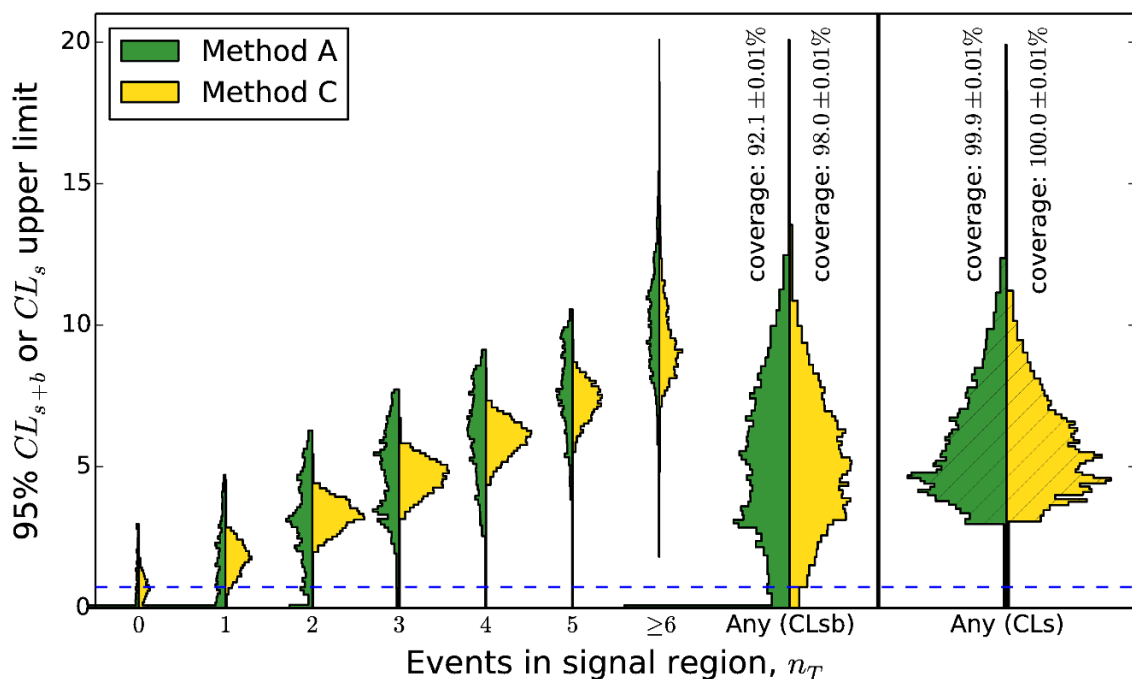


**Figure 3.** For the *simple scenario* described in the text (two leptons in two categories) in which computations using Method B remain tractable, the estimated fake rates for each of 19000 independent toy datasets are shown as a function of  $n_T$ , comparing Method A and Method B with box plots. The fake rate from Method C is by definition the same as that from Method B. The box plots indicate the median and lower & upper quartiles with the box, while the whiskers extend to most extreme datum within  $1.5 \times$  inter-quartile range of the nearest quartile; this corresponds to the  $k = 1.5$  case as described in [42]. Black dots are used to mark data points outside the range of the whiskers. The dashed blue line marks the true value of  $\nu_{TF} = 1.94$ , and the red line delimits the unphysical  $\nu_{TF} < 0$  region.

$CL_{s+b}$  limit are significantly more pronounced in Method A than in Method C, as can be seen when the limits are separated by  $n_T$ , as also included in figure 4.

## 4 Conclusions

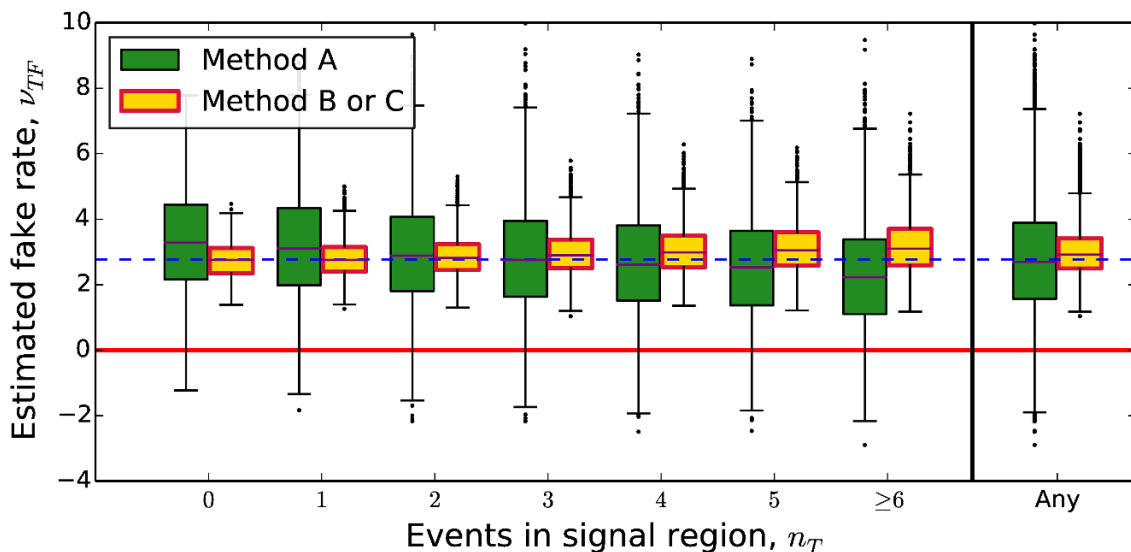
We have described the matrix method, used in many ATLAS and CMS analyses to estimate fake leptonic backgrounds, more completely than we have seen elsewhere. We have shown that it (Method A) produces a MLE fake rate under a restrictive set of conditions, but that these are rarely met in practice. We have shown that it has a number of undesired properties which result from its heuristic definition: (i) it can give physically meaningless results (predict negative fake rates), (ii) its fake rate estimates show an undesired bias as a function of the number of tight events in the signal region (as seen by the slope observed in the run of green boxes of figures 3 and 5), and (iii) the limits it sets on fake rates are significantly more variable than those from better methods (as seen by the increased vertical extent of the Method A histograms in figures 2 and 4 compared to those of Method B and Method C). We noted that, within the constraints of the frequentist profile-likelihood based framework considered, one cannot hope to constrain fake rates much better than Method B. However, we saw that the computational overheads of Method B precluded its



**Figure 4.** In the *harder scenario* described in the text (two leptons, eight categories) computations using Method B are no longer tractable. This figure therefore compares the only two usable Methods (A and C) in this more challenging case. The 95%  $CL_{s+b}$  and  $CL_s$  upper limits on the rate of  $T$  signal events for each of 19000 independent toy datasets are histogrammed. For each ‘column’, histograms are made for each of Methods A and C and plotted back-to-back. The  $CL_{s+b}$  results are further divided into bins of observed  $n_T$ ; in all cases the area of each histogram is proportional to the number of toy datasets used to create it. The dashed blue line indicates the true signal production rate,  $\nu_{TR} = 0.748$ . The coverage of the observed limits of this truth rate are noted for the overall  $CL_{s+b}$  and  $CL_s$  results.

use in all but the simplest of cases. Finally, we showed that it was possible to find a third approach, Method C, which is computationally of similar complexity to Method A and, though to some extent also heuristic in its definition, nonetheless reproduces much more closely the fake rate estimates of Method B. Method C, in contrast to the matrix method: (i) gives only physically meaningful results (predicts positive fake rates), (ii) its fake rate estimates are unbiased as a function of the number of tight events in the signal region (as seen by the lack of slope across the yellow boxes of figures 3 and 5), and (iii) the limits it sets on fake rates are significantly less variable than those of Method A while being very close to those seen in the optimal Method B (again see figures 2 and 4). The improvements seen in Method C over Method A are particularly notable in signal regions having few events.

A possible advantage of Method B and Method C not explored in this paper is afforded by their ability, if desired, to encapsulate background processes which can contribute by both real and fake events e.g. in different decay modes by use of different parameter trees.



**Figure 5.** For the *harder scenario* described in the text (two leptons, eight categories) the fake rates estimated by Methods A and C for each of 19000 independent toy datasets are shown as a function of  $n_T$  in the form of box plots, similarly to figure 3. The dashed blue line marks the true value of  $\nu_{TF} = 2.77$ , and the red line delimits the unphysical  $\nu_{TF} < 0$  region.

Further to this, it may be found that measurements in additional regions could constrain some of these parameters, much like the efficiencies are already constrained. Whilst speculative, further research in this area has the potential to provide even greater benefits over the original matrix method.

## Acknowledgments

The authors acknowledge funding and support from the Science and Technology Facilities Research Council (STFC) and Peterhouse, and wish to thank Kyle Cranmer for useful discussions relating to the possibility of use of methods similar to Method B.

## A Origin of matrix method approximation

It was stated earlier that, under appropriate conditions,  $\hat{n}_R$  and  $\hat{n}_F$  are maximum likelihood estimators for  $\langle n_R \rangle \equiv \mathbb{E}[n_R|n_T, n_L]$  and  $\langle n_F \rangle \equiv \mathbb{E}[n_F|n_T, n_L]$ . This result, together with its limitations, is now presented.

### A.1 Single lepton and single category

We shall firstly demonstrate this approximation in the simplified case with just a single lepton and category. The corresponding fully general derivation follows in the next section, but the logic it follows is largely the same.

When considering a likelihood as a product of Poisson terms as in equation (2.13), and neglecting the Gaussian terms involving the efficiencies, the negative log likelihood for the term arising from the background component will be

$$-\ln \mathcal{L} = \sum_{\mathcal{T} \in \{T, L\}} (\nu_{\mathcal{T}} - n_{\mathcal{T}} \ln \nu_{\mathcal{T}}). \quad (\text{A.1})$$

Here we sum over two constraints, one from tight events and the other from loose. The means of Poisson distributions are denoted by  $\nu_{\mathcal{T}}$ . From equations (2.2) and (2.10), one finds

$$\nu_{\mathcal{T}} = \sum_{\mathcal{R} \in \{R, F\}} \phi_{\mathcal{T}\mathcal{R}} \nu_{\mathcal{R}}, \quad (\text{A.2})$$

where  $\nu_{\mathcal{R}}$  denotes the means of the Poisson distributions from which the numbers of real and fake events are drawn, and  $\phi$  is the matrix of efficiencies with indices  $\mathcal{T}$  and  $\mathcal{R}$  referring to tight/loose and real/fake properties respectively.

One can now differentiate equation (A.1) with respect to  $\nu_{\mathcal{R}}, \forall \mathcal{R} \in \{R, F\}$ , using the identity in equation (A.2), and find the MLE values for the rates, denoted  $\nu_{\mathcal{T}, \text{MLE}}$ . In order to locate the minimum of the negative log likelihood, one sets these derivatives to 0, yielding

$$\sum_{\mathcal{T} \in \{T, L\}} \left(1 - \frac{n_{\mathcal{T}}}{\nu_{\mathcal{T}, \text{MLE}}}\right) \phi_{\mathcal{T}\mathcal{R}} = 0, \quad \forall \mathcal{R} \in \{R, F\}. \quad (\text{A.3})$$

These are satisfied if  $\nu_{\mathcal{T}, \text{MLE}} = n_{\mathcal{T}} \forall \mathcal{T} \in \{T, L\}$ , the result of which being that, upon inversion, equation (A.2) will look like

$$\nu_{\mathcal{R}, \text{MLE}} = \sum_{\mathcal{T} \in \{T, L\}} \phi_{\mathcal{R}\mathcal{T}}^{-1} n_{\mathcal{T}} = \hat{n}_{\mathcal{R}}, \quad (\text{A.4})$$

analogously to equation (2.3).

Whilst this is a valid operation for the problem as stated above, it should be noted that the minimum of  $-\ln \mathcal{L}$  is represented by equation (A.3) *only* when the components of  $\nu_{\mathcal{R}, \text{MLE}}$  are  $> 0$ .

To conclude, equation (A.4) shows how the matrix method estimator  $\hat{n}_{\mathcal{R}}$  is identical to the MLE values  $\nu_{\mathcal{R}, \text{MLE}}$  in the simplified single-lepton matrix method, *if* the condition above is met.

## A.2 Multiple leptons and multiple categories

When considering a likelihood as a product of Poisson terms as in equation (2.13), and neglecting the Gaussian terms involving the efficiencies, the negative log likelihood for the term arising from the background component will be

$$-\ln \mathcal{L} = \sum_{\omega, \beta} (\nu_{\mathcal{T}\beta}^{\omega} - n_{\mathcal{T}\beta}^{\omega} \ln \nu_{\mathcal{T}\beta}^{\omega}), \quad (\text{A.5})$$



Object category	Signal			Background			$\varepsilon_r$	$\varepsilon_f$
	$\nu_{\bar{L}}$	$\beta$	$\pi$	$\nu_{\bar{L}}$	$\beta$	$\pi$		
$\omega_1$	0.01	0.6	0	0.99	0.6	1	0.8	0.1
$\omega_2$	—	0.4	0	—	0.4	1	0.9	0.2

**Table 2.** Parameters controlling the simple scenario with exactly two leptons, and two categories for each lepton. The parameters are as described in figure 1, however  $\alpha_2 = 1$  and  $\alpha_i = 0 \ \forall i \neq 2$ . The overall production rate of events is  $\nu_{\bar{L}}$ , each one of which is filtered through the decision tree. Components marked with a ‘—’ are not applicable in the context.

where for a set of  $m$  leptons the categories and tight/looseness information are compacted into vectors  $\omega$  and  $\beta$  of length  $m$  respectively. Note also that the means of Poisson distributions are denoted in the general notation by e.g.  $\nu_{\mathcal{T}}$ . From equations (2.2) and (2.10), one finds

$$\nu_{\mathcal{T}\beta}^{\omega} = \sum_{\alpha'} \Phi_{\omega\beta}^{\alpha'} \nu_{\mathcal{R}\alpha'}^{\omega}, \quad \Phi_{\omega\beta}^{\alpha'} = \phi_{\omega_{i_1}\beta_1}^{\alpha'_1} \phi_{\omega_{i_2}\beta_2}^{\alpha'_2} \dots, \tag{A.6}$$

where  $\alpha'$  is a vector representing whether each lepton is real or fake.

One can now differentiate equation (A.5) with respect to  $\nu_{\mathcal{R}\alpha}^{\omega}, \forall \omega, \alpha$ , using the identity in equation (A.6), and find the MLE values for the rates, denoted e.g.  $\nu_{\mathcal{T},\text{MLE}}^{\omega}$ . In order to locate the minimum of the negative log likelihood, one sets all these derivatives to 0, yielding

$$\sum_{\beta'} \left( 1 - \frac{n_{\mathcal{T}\beta'}^{\omega}}{\nu_{\mathcal{T},\text{MLE}}^{\omega}} \right) \Phi_{\omega\alpha}^{\beta'} = 0, \quad \forall \omega, \alpha. \tag{A.7}$$

These are satisfied if  $\nu_{\mathcal{T},\text{MLE}}^{\omega} = n_{\mathcal{T}\beta'}^{\omega}, \forall \beta'$ , the result of which being that, upon inversion, equation (A.6) will look like

$$\nu_{\mathcal{R},\text{MLE}}^{\omega} = \sum_{\beta'} \Phi_{\omega\beta'}^{-1} n_{\mathcal{T}\beta'}^{\omega} = \hat{n}_{\mathcal{R}\alpha}^{\omega}, \tag{A.8}$$

analogously to equation (2.3).

Whilst this is a valid operation for the problem as stated above, it should be noted that the minimum of  $-\ln \mathcal{L}$  is represented by equation (A.7) *only* when the components of  $\nu_{\mathcal{R},\text{MLE}}^{\omega}$  are  $> 0$ . Incidentally, is also only useful in the case where the components of  $\nu_{\mathcal{R},\text{MLE}}^{\omega}$  are readily assigned to either signal plus other ‘real’ backgrounds (those typically estimated from MC samples) and the fake background.

To conclude, equation (A.8) shows how the matrix method estimator  $\hat{n}_{\mathcal{R}}$  is identical to the MLE values  $\nu_{\mathcal{R},\text{MLE}}$  in the fully generalised matrix method, *if* the conditions above are met.

## B Tables of parameters

Parameters used to configure the toy generator may be found in tables 2 and 3.

Object category	Signal			Background			$\varepsilon_r$	$\varepsilon_f$
	$\nu_{\bar{L}}$	$\beta$	$\pi$	$\nu_{\bar{L}}$	$\beta$	$\pi$		
$\omega_1$	0.01	0.086	0	0.99	0.184	1	0.8	0.1
$\omega_2$	—	0.143	0	—	0.008	1	0.8	0.2
$\omega_3$	—	0.110	0	—	0.182	1	0.8	0.1
$\omega_4$	—	0.010	0	—	0.123	1	0.8	0.3
$\omega_5$	—	0.092	0	—	0.102	1	0.9	0.2
$\omega_6$	—	0.284	0	—	0.081	1	0.9	0.1
$\omega_7$	—	0.245	0	—	0.106	1	0.9	0.4
$\omega_8$	—	0.030	0	—	0.214	1	0.9	0.1

**Table 3.** Parameters controlling the simple scenario with exactly two leptons, and eight categories for each lepton. The parameters are as described in figure 1, however  $\alpha_2 = 1$  and  $\alpha_i = 0 \forall i \neq 2$ . The overall production rate of events is  $\nu_{\bar{L}}$ , each one of which is filtered through the decision tree. Components marked with a ‘—’ are not applicable in the context.

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

## References

- [1] ATLAS collaboration, *Search for strong production of supersymmetric particles in final states with missing transverse momentum and at least three b-jets using 20.1 fb<sup>-1</sup> of pp collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector*, [ATLAS-CONF-2013-061](#), CERN, Geneva Switzerland (2013) [[INSPIRE](#)].
- [2] ATLAS collaboration, *Background studies for top-pair production in lepton plus jets final states in  $\sqrt{s} = 7$  TeV ATLAS data*, [ATLAS-CONF-2010-087](#), CERN, Geneva Switzerland (2010) [[INSPIRE](#)].
- [3] ATLAS collaboration, *Search for supersymmetry using final states with one lepton, jets and missing transverse momentum with the ATLAS detector in  $\sqrt{s} = 7$  TeV pp*, *Phys. Rev. Lett.* **106** (2011) 131802 [[arXiv:1102.2357](#)] [[INSPIRE](#)].
- [4] ATLAS collaboration, *Search for an excess of events with an identical flavour lepton pair and significant missing transverse momentum in  $\sqrt{s} = 7$  TeV proton-proton collisions with the ATLAS detector*, *Eur. Phys. J. C* **71** (2011) 1647 [[arXiv:1103.6208](#)] [[INSPIRE](#)].
- [5] ATLAS collaboration, *Search for supersymmetric particles in events with lepton pairs and large missing transverse momentum in  $\sqrt{s} = 7$  TeV proton-proton collisions with the ATLAS experiment*, *Eur. Phys. J. C* **71** (2011) 1682 [[arXiv:1103.6214](#)] [[INSPIRE](#)].
- [6] ATLAS collaboration, *Search for pair production of first or second generation leptoquarks in proton-proton collisions at  $\sqrt{s} = 7$  TeV using the ATLAS detector at the LHC*, *Phys. Rev. D* **83** (2011) 112006 [[arXiv:1104.4481](#)] [[INSPIRE](#)].

- [7] ATLAS collaboration, *Measurement of the cross section for the production of a  $W$  boson in association with  $b^-$  jets in  $pp$  collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector*, *Phys. Lett. B* **707** (2012) 418 [[arXiv:1109.1470](#)] [[INSPIRE](#)].
- [8] ATLAS collaboration, *Search for anomalous production of prompt like-sign muon pairs and constraints on physics beyond the standard model with the ATLAS detector*, *Phys. Rev. D* **85** (2012) 032004 [[arXiv:1201.1091](#)] [[INSPIRE](#)].
- [9] ATLAS collaboration, *Search for pair-produced heavy quarks decaying to  $Wq$  in the two-lepton channel at  $\sqrt{s} = 7$  TeV with the ATLAS detector*, *Phys. Rev. D* **86** (2012) 012007 [[arXiv:1202.3389](#)] [[INSPIRE](#)].
- [10] ATLAS collaboration, *Observation of spin correlation in  $t\bar{t}$  events from  $pp$  collisions at  $\sqrt{s} = 7$  TeV using the ATLAS detector*, *Phys. Rev. Lett.* **108** (2012) 212001 [[arXiv:1203.4081](#)] [[INSPIRE](#)].
- [11] ATLAS collaboration, *A search for  $t\bar{t}$  resonances with the ATLAS detector in  $2.05\text{ fb}^{-1}$  of proton-proton collisions at  $\sqrt{s} = 7$  TeV*, *Eur. Phys. J. C* **72** (2012) 2083 [[arXiv:1205.5371](#)] [[INSPIRE](#)].
- [12] ATLAS collaboration, *A search for  $t\bar{t}$  resonances in lepton+jets events with highly boosted top quarks collected in  $pp$  collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector*, *JHEP* **09** (2012) 041 [[arXiv:1207.2409](#)] [[INSPIRE](#)].
- [13] ATLAS collaboration, *ATLAS search for a heavy gauge boson decaying to a charged lepton and a neutrino in  $pp$  collisions at  $\sqrt{s} = 7$  TeV*, *Eur. Phys. J. C* **72** (2012) 2241 [[arXiv:1209.4446](#)] [[INSPIRE](#)].
- [14] ATLAS collaboration, *Search for resonant top plus jet production in  $t\bar{t} + \text{jets}$  events with the ATLAS detector in  $pp$  collisions at  $\sqrt{s} = 7$  TeV*, *Phys. Rev. D* **86** (2012) 091103 [[arXiv:1209.6593](#)] [[INSPIRE](#)].
- [15] ATLAS collaboration, *Search for supersymmetry in events with large missing transverse momentum, jets and at least one tau lepton in 7 TeV proton-proton collision data with the ATLAS detector*, *Eur. Phys. J. C* **72** (2012) 2215 [[arXiv:1210.1314](#)] [[INSPIRE](#)].
- [16] ATLAS collaboration, *Measurement of  $ZZ$  production in  $pp$  collisions at  $\sqrt{s} = 7$  TeV and limits on anomalous  $ZZZ$  and  $ZZ\gamma$  couplings with the ATLAS detector*, *JHEP* **03** (2013) 128 [[arXiv:1211.6096](#)] [[INSPIRE](#)].
- [17] ATLAS collaboration, *Multi-channel search for squarks and gluinos in  $\sqrt{s} = 7$  TeV  $pp$  collisions with the ATLAS detector*, *Eur. Phys. J. C* **73** (2013) 2362 [[arXiv:1212.6149](#)] [[INSPIRE](#)].
- [18] ATLAS collaboration, *Search for  $t\bar{t}$  resonances in the lepton plus jets final state with ATLAS using  $4.7\text{ fb}^{-1}$  of  $pp$  collisions at  $\sqrt{s} = 7$  TeV*, *Phys. Rev. D* **88** (2013) 012004 [[arXiv:1305.2756](#)] [[INSPIRE](#)].
- [19] ATLAS collaboration, *Measurement of top quark polarization in top-antitop events from proton-proton collisions at  $\sqrt{s} = 7$  TeV using the ATLAS detector*, *Phys. Rev. Lett.* **111** (2013) 232002 [[arXiv:1307.6511](#)] [[INSPIRE](#)].
- [20] ATLAS collaboration, *Measurement of the top quark pair production charge asymmetry in proton-proton collisions at  $\sqrt{s} = 7$  TeV using the ATLAS detector*, *JHEP* **02** (2014) 107 [[arXiv:1311.6724](#)] [[INSPIRE](#)].

- [21] ATLAS collaboration, *Search for a multi-Higgs-boson cascade in  $W^+W^-b\bar{b}$  events with the ATLAS detector in  $pp$  collisions at  $\sqrt{s} = 8$  TeV*, *Phys. Rev. D* **89** (2014) 032002 [[arXiv:1312.1956](#)] [[INSPIRE](#)].
- [22] ATLAS collaboration, *Measurement of the production of a  $W$  boson in association with a charm quark in  $pp$  collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector*, *JHEP* **05** (2014) 068 [[arXiv:1402.6263](#)] [[INSPIRE](#)].
- [23] ATLAS collaboration, *Search for direct top-squark pair production in final states with two leptons in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector*, *JHEP* **06** (2014) 124 [[arXiv:1403.4853](#)] [[INSPIRE](#)].
- [24] ATLAS collaboration, *Search for direct top squark pair production in events with a  $Z$  boson,  $b$ -jets and missing transverse momentum in  $\sqrt{s} = 8$  TeV  $pp$  collisions with the ATLAS detector*, *Eur. Phys. J. C* **74** (2014) 2883 [[arXiv:1403.5222](#)] [[INSPIRE](#)].
- [25] ATLAS collaboration, *Search for direct production of charginos, neutralinos and sleptons in final states with two leptons and missing transverse momentum in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector*, *JHEP* **05** (2014) 071 [[arXiv:1403.5294](#)] [[INSPIRE](#)].
- [26] ATLAS collaboration, *Search for supersymmetry at  $\sqrt{s} = 8$  TeV in final states with jets and two same-sign leptons or three leptons with the ATLAS detector*, *JHEP* **06** (2014) 035 [[arXiv:1404.2500](#)] [[INSPIRE](#)].
- [27] ATLAS collaboration, *Search for microscopic black holes and string balls in final states with leptons and jets with the ATLAS detector at  $\sqrt{s} = 8$  TeV*, *JHEP* **08** (2014) 103 [[arXiv:1405.4254](#)] [[INSPIRE](#)].
- [28] ATLAS collaboration, *Search for top squark pair production in final states with one isolated lepton, jets and missing transverse momentum in  $\sqrt{s} = 8$  TeV  $pp$  collisions with the ATLAS detector*, [arXiv:1407.0583](#) [[INSPIRE](#)].
- [29] ATLAS collaboration, *Search for strong production of supersymmetric particles in final states with missing transverse momentum and at least three  $b$ -jets at  $\sqrt{s} = 8$  TeV proton-proton collisions with the ATLAS detector*, *JHEP* **1410** (2014) 24 [[arXiv:1407.0600](#)] [[INSPIRE](#)].
- [30] ATLAS collaboration, *Search for supersymmetry in events with large missing transverse momentum, jets and at least one tau lepton in  $20\text{fb}^{-1}$  of  $\sqrt{s} = 8$  TeV proton-proton collision data with the ATLAS detector*, *JHEP* **09** (2014) 103 [[arXiv:1407.0603](#)] [[INSPIRE](#)].
- [31] ATLAS collaboration, *Measurement of the  $t\bar{t}$  production cross-section as a function of jet multiplicity and jet transverse momentum in 7 TeV proton-proton collisions with the ATLAS detector*, [arXiv:1407.0891](#) [[INSPIRE](#)].
- [32] CMS collaboration, *Measurement of the top-quark mass in  $t\bar{t}$  events with dilepton final states in  $pp$  collisions at  $\sqrt{s} = 7$  TeV*, *Eur. Phys. J. C* **72** (2012) 2202 [[arXiv:1209.2393](#)] [[INSPIRE](#)].
- [33] CMS collaboration, *Identification of  $b$ -quark jets with the CMS experiment*, *2013 JINST* **8** P04013 [[arXiv:1211.4462](#)] [[INSPIRE](#)].
- [34] CMS collaboration, *Search for a Higgs boson decaying into a  $b$ -quark pair and produced in association with  $b$  quarks in proton-proton collisions at 7 TeV*, *Phys. Lett. B* **722** (2013) 207 [[arXiv:1302.2892](#)] [[INSPIRE](#)].

- [35] CMS collaboration, *Measurement of the  $t\bar{t}$  production cross section in the dilepton channel in  $pp$  collisions at  $\sqrt{s} = 7$  TeV*, *JHEP* **11** (2012) 067 [[arXiv:1208.2671](#)] [[INSPIRE](#)].
- [36] B. Mistlberger and F. Dulat, *Limit setting procedures and theoretical uncertainties in Higgs boson searches*, [arXiv:1204.3851](#) [[INSPIRE](#)].
- [37] BIPM, IEC, IFCC, ISO, IUPAC, IUPAP and OIML, *Guide to the expression of uncertainty in measurement*, International Organization for Standardization, (1995).
- [38] ATLAS collaboration, K. Cranmer, private communication, (2013).
- [39] F. James and M. Roos, *Minuit: a system for function minimization and analysis of the parameter errors and correlations*, *Comput. Phys. Commun.* **10** (1975) 343 [[INSPIRE](#)].
- [40] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J. C* **71** (2011) 1554 [[arXiv:1007.1727](#)] [[INSPIRE](#)].
- [41] K.S. Cranmer, *Frequentist hypothesis testing with background uncertainty*, in *Statistical problems in particle physics, astrophysics and cosmology*, L. Lyons, R. Mount and R. Reitmeyer eds., World Scientific, Singapore (2003), pg. 261 [*eConf C 030908* (2003) WEMT004] [[physics/0310108](#)] [[INSPIRE](#)].
- [42] M. Frigge, D.C. Hoaglin and B. Iglewicz, *Some implementations of the boxplot*, *Amer. Statist.* **43** (1989) 50.