Physics Letters B 743 (2015) 1–5

Contents lists available at ScienceDirect

Physics Letters B

www.elsevier.com/locate/physletb

Discoveries far from the lamppost with matrix elements and ranking

Dipsikha Debnath, James S. Gainer *, Konstantin T. Matchev

Physics Department, University of Florida, Gainesville, FL 32611, USA

ARTICLE INFO

Article history:
Received 5 December 2014
Received in revised form 17 January 2015
Accepted 7 February 2015
Available online 11 February 2015
Editor: G.F. Giudice

ABSTRACT

The prevalence of null results in searches for new physics at the LHC motivates the effort to make these searches as model-independent as possible. We describe procedures for adapting the Matrix Element Method for situations where the signal hypothesis is not known *a priori*. We also present general and intuitive approaches for performing analyses and presenting results, which involve the flattening of background distributions using likelihood information. The first flattening method involves ranking events by background matrix element, the second involves quantile binning with respect to likelihood (and other) variables, and the third method involves reweighting histograms by the inverse of the background distribution.

1. Introduction

The CERN Large Hadron Collider (LHC) will soon resume operation, probing energies never before accessed with colliders. Ideas about what sort of new physics it will discover often focus on models that resolve the hierarchy problem [1] or provide relic dark matter candidates [2]. There are a great variety of ideas in each category. However, the lack of convincing evidence for new physics at the LHC to date suggests that we may be looking in the wrong places. We therefore consider methods that will allow the discovery of *any* departure from known physics.

The Matrix Element Method (MEM) [3] and similar multivariate analyses [4] have been used with success in LHC experiments. A particularly dramatic example has been the use of the MEM in the four-lepton channel for the discovery of the Higgs boson [5] and the measurement of its properties [6]. Such analyses used variables, such as MELA KD [7] or MEKD [8], that involve the ratio of signal and background matrix elements. Clearly these variables are, therefore, optimized to the appropriate signal and background hypotheses.

It is therefore interesting to ask whether some of the advantages of the MEM can be retained without (as is usually done) assuming a specific signal hypothesis. To this end, we propose the use of the background matrix element alone (or some related quantity) as a kinematic variable. This variable automatically "knows" about important properties of the background, such as spin correlations and kinematic edges. To make such variables more intuitive, we suggest "flattening" the distribution of these variables in various ways.

In what follows, we will briefly review the MEM and highlight the differences in the case where there is no signal hypothesis. We will then describe in detail three useful methods for flattening the distribution, to help make analyses using the background matrix element more intuitive.

2. The matrix element method

According to the Neyman–Pearson lemma [9], the optimal test statistic [17] for comparing hypotheses H_0 and H_1 is provided by the likelihood ratio:

$$R_\Lambda(H_0, H_1) = \frac{\Lambda_{H_0}(\{\mathcal{E}_i\})}{\Lambda_{H_1}(\{\mathcal{E}_i\})}, \tag{1}$$

where Λ_{H_α} is the likelihood for the hypothesis H_α ($\alpha = 1, 2$) as a function of the data, which we assume consists of N events, $\mathcal{E}_i$, $i = 1, 2, \ldots, N$. Each event consists of a set $\left\{\tilde{p}_j^\mu\right\}$ of measured momenta for N_{vis} particles.

$$\mathcal{E}_i \equiv \left\{\tilde{p}_j^\mu\right\}_i, \quad j = 1, 2, \ldots, N_{vis}. \tag{2}$$

In the MEM, the likelihood for a given event (2) is calculated using the expression

* Corresponding author.
E-mail address: jgainer137@gmail.com (J.S. Gainer).

http://dx.doi.org/10.1016/j.physletb.2015.02.020

$$\mathcal{P}(\mathcal{E}_i|H_\alpha) = \frac{1}{\sigma(H_\alpha)} \times \left[\prod_{j=1}^{N_{vis}} \int \frac{d^3 p_j}{(2\pi)^3 2E_j}\right] \times T(\{\tilde{p}_j\},\{p_j\})$$
$$\times \left[\prod_{k=1}^{N_{inv}^{(\alpha)}} \int \frac{d^3 q_k}{(2\pi)^3 2E_k}\right] \sum_{a,b} \frac{f_a(x_1) f_b(x_2)}{2s x_1 x_2}$$
$$\times |\mathcal{M}_{H_\alpha,ab}(\{p_j\},\{q_k\})|^2, \tag{3}$$

where $\mathcal{M}_{H_\alpha,ab}$ is the theoretical matrix element for hypothesis H_α, f_a and f_b are parton distribution functions (pdf) as a function of momentum fractions x_1 and x_2, while $\sigma(H_\alpha)$ is the total cross section after acceptances, efficiencies, etc. Note that the theoretical matrix element $\mathcal{M}_{H_\alpha,ab}$ is a function of the corresponding *true* momenta $\{p_j\}$ of the N_{vis} visible particles, which are mapped onto the set of measured momenta $\{\tilde{p}_j\}$ via the transfer function $T(\{\tilde{p}_j\},\{p_j\})$ which incorporates all detector effects and efficiencies [10]. In general, a given hypothesis H_α for the event will also involve a certain number $N_{inv}^{(\alpha)}$ of invisible particles (such as neutrinos or dark matter candidates), whose momenta q_k cannot be measured.

The likelihood for a set of N events $\{\mathcal{E}_i\}$ is simply the product of the likelihoods for each event:

$$\Lambda_{H_\alpha}(\{\mathcal{E}_i\}) = \prod_i^N \mathcal{P}(\mathcal{E}_i|H_\alpha). \tag{4}$$

Thus the likelihood ratio (1) contains the product of ratios of event-by-event likelihoods described in Eq. (3). Often, the two hypotheses (H_0 and H_1) will involve the same final state, hence factors due to the phase space integrals in Eq. (3) will cancel in the likelihood ratio. We are then left with a ratio of squared matrix elements, possibly weighted (in the case where the hypotheses involve different initial state partons) by pdfs. These squared matrix elements contain a great deal of information about the process, including the pole structure, spin correlations, etc. While the implementation of an analysis using the likelihood ratio (1) as a test statistic may sometimes be challenging in practice, conceptually the implementation is straightforward and the sensitivity is, at least in principle, optimal [7,8].

3. Discovery from background likelihood distributions

The limitation of the MEM is that we must know the signal process in order to calculate the appropriate likelihood. As a result, if we do not know what signal model we are looking for we can no longer consider the likelihood ratio, as we know only one hypothesis, the background. It will still be useful, however, to use the information about the background that is encoded in the matrix element. Therefore, we propose that we consider the event-by-event background likelihood, $\mathcal{P}(\mathcal{E}_i|\text{bg})$, and closely-related expressions, as *variables*. Here $P(\mathcal{E}_i|\text{ bg})$ is either defined following Eq. (3) for the background hypothesis or is a similar variable (such as the background squared matrix element).

As an example, consider the "golden" four lepton channel in which the Higgs boson discovery was made. We can define the variable

$$\Lambda_B = \sum_{i=1}^N \log|\mathcal{M}_{\text{bg}}(\mathcal{E}_i)|^2, \tag{5}$$

where $\mathcal{M}_{\text{bg}}(\mathcal{E}_i)$ is the pdf-weighted squared matrix element for $q\bar{q} \to 4\ell$ as evaluated for the final state momenta of the $2e2\mu$ event, $\mathcal{E}_i$. This quantity can be calculated using MEKD [8] (a package for MEM calculations for the four-lepton final state based on

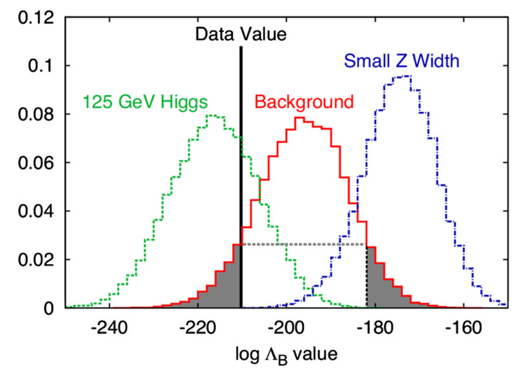

Fig. 1. The unit normalized distribution of Λ_B, defined in Eq. (5) as evaluated for 20-event pseudoexperiments consisting of background $q\bar{q} \to 4\ell$ events (red solid curve), $gg \to H \to 4\ell$ signal events for a 125 GeV Higgs (green dashed curve), and $q\bar{q} \to 4\ell$ events for which the Z boson width has been reduced by a factor of 5 (blue dot-dashed curve). If a particular value of Λ_B, indicated by "Data Value" is observed, the corresponding p-value is given by the area in gray. In this specific case, $p \approx 0.13$.

MadGraph [11]). Fig. 1 shows distributions of Λ_B, evaluated for pseudo-experiments consisting of 20 events, for three different hypotheses: the irreducible $q\bar{q} \to 2e2\mu$ background (red solid curve), gluon fusion production of a 125 GeV Higgs boson that decays to $2e2\mu$ (green dashed curve), and the irreducible $q\bar{q} \to 2e2\mu$ with the Z boson width scaled down by a factor of 5 (blue dot-dashed curve). The fact that the distributions of Λ_B for the different scenarios are well-separated shows that this variable has good sensitivity. In this figure, we also show graphically how a p-value, describing the extent to which actual data is consistent with the background hypothesis, can be obtained numerically for the Λ_B variable.

It is important to note two things:

i) We have chosen to use the pseudo-experiment variable Λ_B, rather than the event by event $|\mathcal{M}|^2$ for convenience of illustration. In general, event-by-event variables will be more sensitive. Many statistical tests can be employed to test whether the values of $|\mathcal{M}|^2$ obtained for observed events agree with the background distribution of this quantity. Popular choices include Fisher's exact test [12], the Kolmogorov–Smirnov test [13], and even the classic χ^2 test [14].
ii) Unlike the likelihood ratio, the numerical value of variables like Λ_B and $|\mathcal{M}|^2$ does not have a direct statistical interpretation. We have highlighted this fact in Fig. 1 by selecting one hypothesis for which $|\Lambda_B|_S > |\Lambda_B|_B$ (blue dot-dashed curve), and one for which $|\Lambda_B|_S < |\Lambda_B|_B$ (green dashed curve). (Here $|\Lambda_B|_{S(B)}$ is the average value of Λ_B for pseudo-experiments of signal (background) events.)

4. How to flatten background distributions: examples

The main point of this letter is that the procedure described above allows one to exclude the background hypothesis in the presence of an *unknown* signal. In other words, one can confidently look for new physics models "away from the lamppost", i.e., models which no theorist has yet thought of. While, in principle, any variable can be used to test the background hypothesis in this way, the use of a variable based on the background likelihood should additionally optimize the sensitivity of such searches.

We now present some related methods, which allow the "non-backgroundness" of some potential signal to be shown in a clear and intuitive way. These methods also have the benefit that they generalize to any possible channel, so results and sensitivity in various channels can easily be compared. We emphasize that the

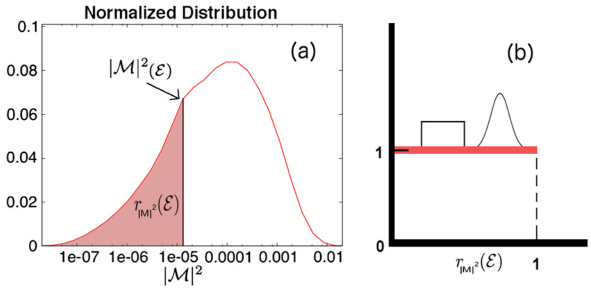

Fig. 2. Panel (a) shows how the distribution of background MEKD for background events is used to create a "ranking variable". In panel (b) we show that the background distribution with respect to this ranking variable is flat, while for other processes the distribution of background ranking variable is not flat.

motivation for the flattening procedures is to facilitate the visualization of excess and deficits in a model and channel-independent way, not to enhance statistical significance. Flattening approaches have been used before [15], though their use is underappreciated, which motivates our relatively detailed description of their use below.

1. Flattening with ranking. In this approach, one takes the normalized distribution, $\frac{dN}{d\xi}$, for some kinematic variable, ξ, and defines a "ranking" variable,

$$r(\xi) = \int_{-\infty}^{\xi} \frac{dN}{d\xi'} d\xi'. \tag{6}$$

We note that $r(\xi)$ is the cumulative distribution function for the background with respect to the variable ξ. We can now evaluate the ranking r_ξ of any given event $\mathcal{E}$ by defining

$$r_\xi(\mathcal{E}) = r(\xi(\mathcal{E})), \tag{7}$$

that is, the value of the ranking variable for a given event, $\mathcal{E}$, is the value found from Eq. (6) for the value of the kinematic variable ξ obtained for the event. The connection between our ranking variable $r_\xi(\mathcal{E})$ and the background ξ distribution is shown pictorially in Fig. 2(a). The figure also illustrates the physical meaning of $r_\xi(\mathcal{E})$ — it is the fraction of background events $\mathcal{E}'$ in which $\xi(\mathcal{E}') < \xi(\mathcal{E})$. If we then consider the normalized distribution of the background with respect to r_ξ, we find that

$$\frac{dN}{dr_\xi} = 1, \tag{8}$$

hence the distribution of this variable for background events is flat, as is shown in Fig. 2(b). This procedure of course works for any kinematic variable, ξ, we especially recommend using it with the sensitive matrix-element-based variables advocated above. Thus, one obtains a sensitive variable for which the background distribution is flat, while the distribution of signal events is characterized by departures from flatness, as is also shown in Fig. 2(b).

We note in passing that calculating $r(\xi)$ from Monte Carlo (MC) events is quite straightforward. One simply calculates the value of the variable ξ for each of the N events in the MC sample, thus obtaining a list of values $\{\xi_i\}$. The value of $r(\xi)$ is then well-approximated by the fraction of the $\{\xi_i\}$ which are less than ξ, i.e.,

$$r(\xi) \approx \frac{1}{N} \sum_{i=1}^{N} \theta(\xi - \xi_i). \tag{9}$$

This approach should facilitate the experimental implementation of this technique, though we note that care must be taken with respect to the possible effects of systematic uncertainties on this distribution.

2. Flattening with quantile bins. An alternate approach is to use the method of quantile bins [18]. If we are only considering one variable, ξ, this approach consists of finding $n+1$ ξ values: $\xi_1, \xi_2, \ldots, \xi_{n+1}$, such that

$$\int_{\xi_i}^{\xi_{i+1}} \frac{dN}{d\xi} d\xi = 1/n, \quad i = 1, \ldots, n, \tag{10}$$

i.e., the integral of the distribution is equal in each bin. This procedure can be extended to the case where there are several variables ξ_i, where again we demand that the integral of the distribution be the same in each bin. For example, in two dimensions, we must choose $n+1$ values of ξ_1: $\xi_{1,1}, \xi_{1,2}, \ldots, \xi_{1,n+1}$ and $n+1$ values of ξ_2: $\xi_{2,1}, \xi_{2,2}, \ldots, \xi_{2,n+1}$, such that

$$\int_{\xi_{1,i}}^{\xi_{1,i+1}} \int_{\xi_{2,j}}^{\xi_{2,j+1}} \frac{d^2N}{d\xi_1 d\xi_2} d\xi_1 d\xi_2 = \frac{1}{n^2} \tag{11}$$

for any i and j. This procedure allows us to consider additional kinematic variables in addition to a likelihood-based variable. Examples of this are shown in Figs. 3 and 4, in which we consider the distribution of four-lepton events at the 8 TeV LHC in terms of the four-lepton invariant mass, $m_{4\ell}$, and the background MEKD value.

In Fig. 3 we show the results of an example experiment where we have formed quantile bins in $m_{4\ell}$ and $|\mathcal{M}|^2$, assuming the background hypothesis. We then plot the number of events in each quantile bin either from 150 background $q\bar{q} \to 2e2\mu$ events (panels in the top row), or 75 125-GeV Higgs signal and 75 background events (panels in the bottom row). The panels in the left column are for one 150 event pseudo-experiment, while the panels in the right column are for the average of 400 such pseudo-experiments. Fig. 4 illustrates the same concept using scatter plots. Here the ratio of signal to background events has been changed from 1:1 (which is realistic for 125 GeV $H \to 4\ell$ signal and the $q\bar{q} \to 4\ell$ background) to the much more challenging 1:3. Nevertheless, the presence of new signal can still be inferred from the anomalous clustering of points. Note that departures from uniform density are easier to interpret in the scatter plot in panel (b), which utilizes ranking variables.

3. Flattening with respect to all the variables. An extreme case of flattening the background distribution with respect to kinematic variables occurs when we consider a complete set of kinematic variables for some process. We can, of course, calculate the boundaries of these bins with Monte Carlo. However, in the limit where we have a good analytic, or at least numerical understanding of the background, we can perform a flattening using the background distribution.

Specifically, if the background (after detector simulation, etc.) is described by the differential distribution $d^nN/d\boldsymbol{\xi}$, then if we weight each background event by $1/(d^nN/d\boldsymbol{\xi})$, we will end up with a distribution that is flat in the full n-dimensional space of values. If we weight data events according to this procedure, a signal will show up as deviations from flatness. This procedure is demonstrated in Fig. 5. The image in the top left represents our background pdf. If we generate "events" (i.e., pixels) according to this pdf, but weigh the corresponding 2D histogram by the reciprocal of the pdf, then we obtain an essentially flat distribution, shown in the top right corner. We now consider the bottom left image, where some "signal" (American football and flying saucers)

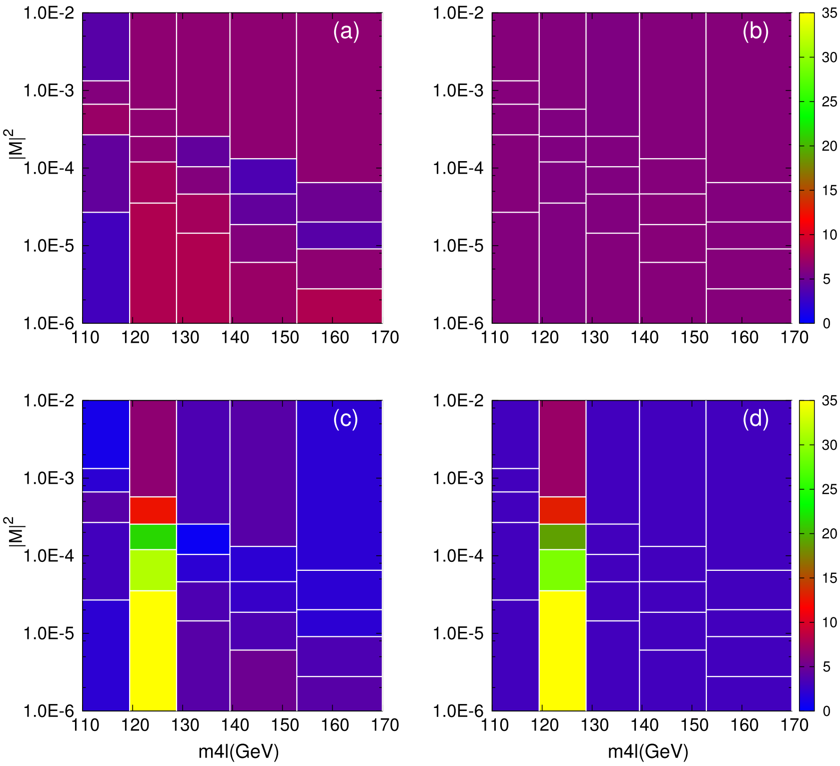

Fig. 3. Quantile bins in $m_{4\ell}$ (x-axis) and $|\mathcal{M}|^2$ (y-axis) we constructed using the background ($q\bar{q} \to 2e2\mu$) distribution. We then plot the number of events in each quantile bin either from 150 background $q\bar{q} \to 2e2\mu$ events (panels in the top row), or 75 125-GeV Higgs signal and 75 background events (panels in the bottom row). The panels in the left column are for one 150 event pseudo-experiment, while the panels in the right column are for the average of 400 pseudo-experiments.

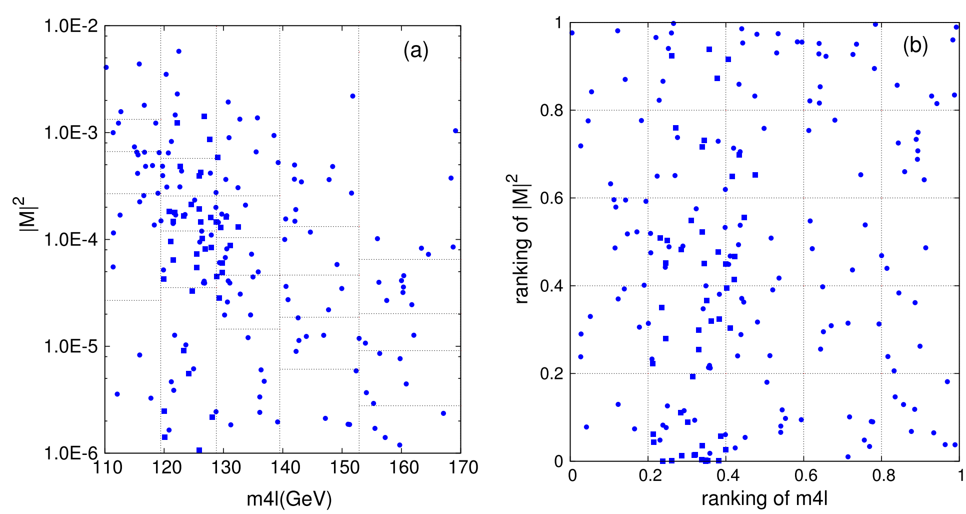

Fig. 4. Simulated data consisting of 50 125-GeV Higgs $gg \to H \to 4\ell$ events and 150 $q\bar{q} \to 4\ell$ events. In panel (a), the four-lepton invariant mass and pdf-weighted background squared matrix element have been plotted, while in panel (b) the ranking variable corresponding to these quantities, as defined in Eq. (6), is plotted. In each case, the dotted line mesh represents the quantile bin boundaries.

have been added to the background. If events are generated according to this pdf, but weighted according to the reciprocal of the background pdf, we obtain the bottom right image, in which background features have been flattened, but signal features remain distinct.

5. Conclusions

We have presented methods, which utilize variables based on the squared matrix element, to search for new physics signals at the LHC in a model independent way. These approaches allow for model-independent exclusions of the standard model in the presence of arbitrary, unspecified, new physics. We look forward to the utilization of such methods in the upcoming Run 2 at the LHC.

Acknowledgements

J.G. and K.M. would like to thank their CMS colleagues for useful discussions. All of the authors would also like to thank D. Kim and T. LeCompte for stimulating conversations. Work supported in part by U.S. Department of Energy Grant DE-FG02-97ER41029 (grant number is DE-SC0010296).

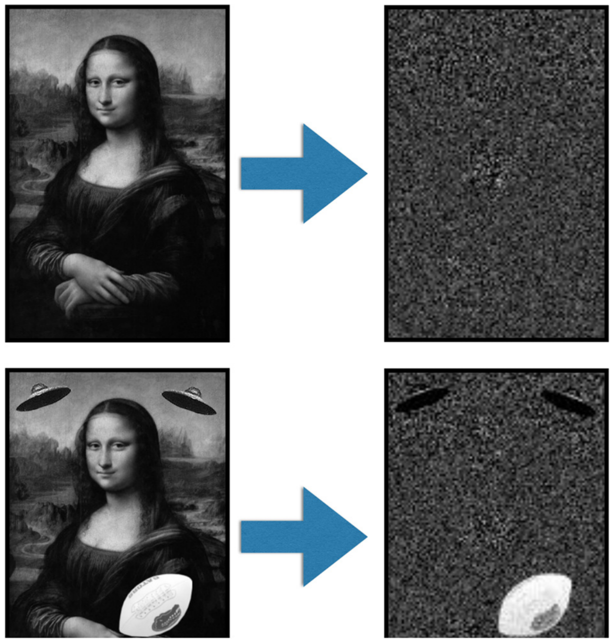

Fig. 5. A demonstration of filling histograms with the reciprocal of the pdf for the case of background only (top row) and in the presence of both signal and background (bottom row).

References

[1] L. Susskind, Phys. Rev. D 20 (1979) 2619;
G. 't Hooft, NATO ASI Ser., Ser. B: Phys. 59 (1980) 135.
[2] J.R. Ellis, J.S. Hagelin, D.V. Nanopoulos, K.A. Olive, M. Srednicki, Nucl. Phys. B 238 (1984) 453;
G. Jungman, M. Kamionkowski, K. Griest, Phys. Rep. 267 (1996) 195, arXiv:hep-ph/9506380;
G. Bertone, D. Hooper, J. Silk, Phys. Rep. 405 (2005) 279, arXiv:hep-ph/0404175.
[3] K. Kondo, J. Phys. Soc. Jpn. 57 (1988) 4126;
K. Kondo, J. Phys. Soc. Jpn. 60 (1991) 836;
K. Kondo, T. Chikamatsu, S.H. Kim, J. Phys. Soc. Jpn. 62 (1993) 1177;
R.H. Dalitz, G.R. Goldstein, Phys. Rev. D 45 (1992) 1531;
B. Abbott, et al., D0 Collaboration, Phys. Rev. D 60 (1999) 052001, arXiv:hep-ex/9808029;
J.C. Estrada Vigil, FERMILAB-THESIS-2001-07;
M.F. Canelli, UMI-31-14921;
V.M. Abazov, et al., D0 Collaboration, Nature 429 (2004) 638, arXiv:hep-ex/0406031;
J.S. Gainer, J. Lykken, K.T. Matchev, S. Mrenna, M. Park, arXiv:1307.3546 [hep-ph].
[4] P.C. Bhat, Annu. Rev. Nucl. Part. Sci. 61 (2011) 281.
[5] S. Chatrchyan, et al., CMS Collaboration, Phys. Lett. B 716 (2012) 30, arXiv:1207.7235 [hep-ex].
[6] S. Chatrchyan, et al., CMS Collaboration, Phys. Rev. Lett. 110 (2013) 081803, arXiv:1212.6639 [hep-ex];
G. Aad, et al., ATLAS Collaboration, Phys. Lett. B 726 (2013) 120, arXiv:1307.1432 [hep-ex];
S. Chatrchyan, et al., CMS Collaboration, arXiv:1312.5353 [hep-ex].
[7] Y. Gao, A.V. Gritsan, Z. Guo, K. Melnikov, M. Schulze, N.V. Tran, Phys. Rev. D 81 (2010) 075022, arXiv:1001.3396 [hep-ph];
A. De Rujula, J. Lykken, M. Pierini, C. Rogan, M. Spiropulu, Phys. Rev. D 82 (2010) 013003, arXiv:1001.5300 [hep-ph];
S. Bolognesi, Y. Gao, A.V. Gritsan, K. Melnikov, M. Schulze, N.V. Tran, A. Whitbeck, Phys. Rev. D 86 (2012) 095031, arXiv:1208.4018 [hep-ph].
[8] P. Avery, D. Bourilkov, M. Chen, T. Cheng, A. Drozdetskiy, J.S. Gainer, A. Korytov, K.T. Matchev, et al., Phys. Rev. D 87 (5) (2013) 055006, arXiv:1210.0896 [hep-ph];
M. Chen, T. Cheng, J.S. Gainer, A. Korytov, K.T. Matchev, P. Milenovic, G. Mitselmakher, M. Park, et al., Phys. Rev. D 89 (2014) 034002, arXiv:1310.1397 [hep-ph].
[9] J. Neyman, E.S. Pearson, Philos. Trans. R. Soc. Lond. A 231 (694–706) (1933) 289–337.
[10] J.S. Gainer, J. Lykken, K.T. Matchev, S. Mrenna, M. Park, J. High Energy Phys. 1410 (2014) 78, arXiv:1404.7129 [hep-ph].
[11] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, T. Stelzer, J. High Energy Phys. 1106 (2011) 128, arXiv:1106.0522 [hep-ph].
[12] R.A. Fisher, J. R. Stat. Soc. 85 (1922) 87–94, http://dx.doi.org/10.2307/2340521, JSTOR 2340521.
[13] A. Kolmogorov, G. Ist. Ital. Attuari 4 (1933) 83–91;
N. Smirnov, Ann. Math. Stat. 19 (1948) 279–281, http://dx.doi.org/10.1214/aoms/1177730256/.
[14] K. Pearson, Philos. Mag. Ser. 5 50 (302) (1900) 157–175, http://dx.doi.org/10.1080/14786440009463897.
[15] B. Aubert, et al., BaBar Collaboration, Phys. Rev. Lett. 91 (2003) 241801, arXiv:hep-ex/0308012.
[16] N. Arkani-Hamed, G.L. Kane, J. Thaler, L.-T. Wang, J. High Energy Phys. 0608 (2006) 070, arXiv:hep-ph/0512190.
[17] We note that optimality is guaranteed for "simple" hypotheses; the generally-necessary inclusion of nuisance parameters which must be integrated over complicates the situation somewhat.
[18] Quantile bins have previously been employed in studies of the LHC inverse problem [16].