A new method for ISR jet tagging[*]

Andrés Felipe García Albarracín[†] and Juan Carlos Sanabria Arenas[‡]
Universidad de los Andes

(XYXYX Collaboration)

(Dated: May 27, 2015)

An article usually includes an abstract, a concise summary of the work covered at length in the main body of the article.

Usage: Secondary publications and information retrieval purposes.

PACS numbers: May be entered using the `\pacs{#1}` command.

Structure: You may use the `description` environment to structure your abstract; use the optional argument of the `\item` command to give the category of each item.

I. INTRODUCTION

II. THE ISR JET TAGGING METHOD

A. The method

Let's suppose that there exists a kinematic variable y that distinguishes between ISR jets and Non ISR jets. The information of such variable is known by means of the distribution functions for each type of jet (f^{ISR}, $f^{Non\ ISR}$). Therefore, if a measurement of the variable y for a particular jet is y_0, then $f^{ISR}(y_0)$ and $f^{Non\ ISR}(y_0)$ are known, as it is presented in Fig. 1.

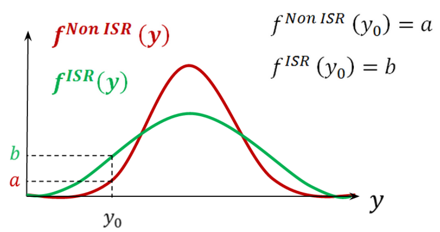

FIG. 1. Probability distributions of a variable that distinguishes between ISR and Non ISR jets

The difference between both distributions could be used to write the probability of such jet being ISR or not. In fact, the probability of being ISR should be proportional to the ISR distribution function at the measurement. Likewise, the probability of being non ISR should be proportional to the Non ISR distribution function:

$$P^{ISR}(y_0) \propto f^{ISR}(y_0), \tag{1}$$

$$P^{Non\ ISR}(y_0) \propto f^{Non\ ISR}(y_0). \tag{2}$$

In addition to the information offered by the density functions, another important consideration to take into account is the *apriori* probability of being ISR. If just one jet of the N_{jets} in the event is ISR, the *apriori* probability of any jet being ISR is:

$$P^{ISR}_{apriori}(y_0) = \frac{1}{N_{jets}}, \tag{3}$$

and similarly, the *apriori* probability of any jet being Non ISR is:

$$P^{Non\ ISR}_{apriori}(y_0) = \frac{N_{jets}-1}{N_{jets}}. \tag{4}$$

Combining both assumptions, the probabilities of being ISR and Non ISR could be written as:

$$P^{ISR}(y_0) = \alpha f^{ISR}(y_0)\frac{1}{N_{jets}}, \tag{5}$$

$$P^{Non\ ISR}(y_0) = \alpha f^{Non\ ISR}(y_0)\frac{N_{jets}-1}{N_{jets}}, \tag{6}$$

where α is a constant that results from the normalization of the probabilities:

$$1 = P^{ISR}(y_0) + P^{FSR}(y_0), \tag{7}$$

$$\alpha = \frac{N_{jets}}{f^{ISR}(y_0) + (N_{jets}-1)f^{Non\ ISR}(y_0)}. \tag{8}$$

If there are more than a single variable which differentiate between ISR and Non ISR jets, the previous analysis can be extended easily. In fact, it is enough to replace de single variable probability density functions by multidimensional probability densities. The formulas would take the same form as the probability density distributions are scalar functions, regardless they depend on a single variable y or on a vector $\vec{y}$. Therefore, in a multidimensional case, the formulas would be:

[*] A footnote to the article title
[†] af.garcia1214@uniandes.edu.co
[‡] jsanabri@uniandes.edu.co

$$P^{ISR}(\vec{y_0}) = \alpha f^{ISR}(\vec{y_0})\frac{1}{N_{jets}}, \tag{9}$$

$$P^{Non\ ISR}(\vec{y_0}) = \alpha f^{Non\ ISR}(\vec{y_0})\frac{N_{jets}-1}{N_{jets}}, \tag{10}$$

B. From probability density functions to normalized histograms

As the latter formulas show, the probabilities of each jet depend on the probability density distributions. In practical matters, these functions are replaced by normalized histograms whose entries are collected from simulations where the ISR jet is known.

However, the replacement is just an approximation because a bin of the histogram does not correspond exactly to the value of the probability density function. In fact, the histogram results from an integration of the probability distribution:

$$H(y_i) = \int_{\Omega_i} f(y)dy, \tag{11}$$

where Ω_i is the range of the bin, as presented in Fig. 2.

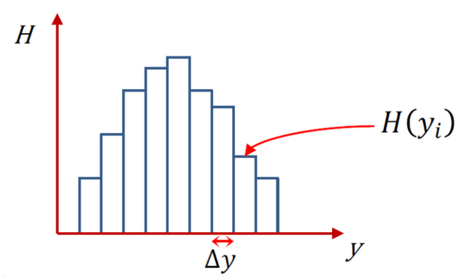

FIG. 2. Shape of a histogram which does not exactly correspond with the probability density function

If the size of the bin is small enough, the expression 11 can be approximated by:

$$H(y_i) \approx f(y_i)\Delta y, \tag{12}$$

Using this approximation, the practical expressions of the probabilities of being ISR or Non ISR are:

$$P^{ISR}(\vec{y_0}) = \alpha H^{ISR}(\vec{y_0})\frac{1}{N_{jets}}, \tag{13}$$

$$P^{Non\ ISR}(\vec{y_0}) = \alpha H^{Non\ ISR}(\vec{y_0})\frac{N_{jets}-1}{N_{jets}}. \tag{14}$$

To sum up, the usage of these formulas implies the necessity of running simulations of several events, identifying theoretically the ISR jet in each event, and filling a N-dimensional histogram for each type of jet (Non ISR and ISR).

C. The Algorithm

Once the method has been prepared by selecting the distinguishing variables and filling the histograms, the algorithm of Fig. 3 is applied for each event. First, each jet in the event is studied and its probabilities of being ISR and Non ISR are determined from its kinematical variables and expressions 9 and 10.

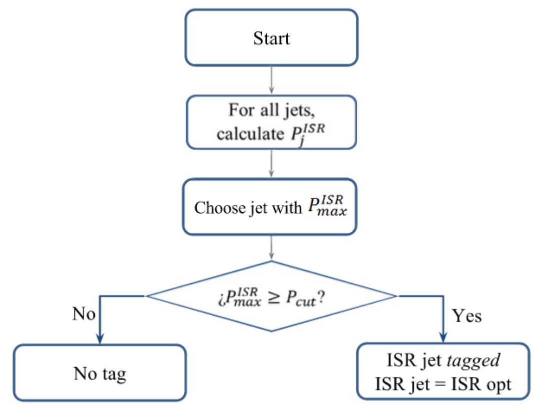

FIG. 3. ISR jet tagging algorithm

Then, the jet with greatest probability of being ISR P^{ISR}_{max} is selected as ISR candidate. Finally, P^{ISR}_{max} is compared to a certain cut P_{cut} in order to ensure that the algorithm is conclusive. For example, if $P^{ISR}_{max} < 1/N_{jets}$, the probability of the ISR candidate is fewer than the *apriori* probability, and therefore no tag should be imposed. The cut is written in terms of a variable k that corresponds to the minimum factor that the probability of the ISR candidate should be greater than the *apriori* probability:

$$P_{cut} = \frac{k}{N_{jets}} \tag{15}$$

This way, the ISR jet is tagged in each event based exclusively on preliminary histograms and simple probability considerations.

III. PREPARATION OF THE METHOD

A. The simulation chain

Overall, the entire project was done using Monte Carlo simulations. With those simulations, the multidimensional density histograms were filled and the events used to test the tagging algorithm were obtained.

Every simulation was performed using the chain MadGraph 5.2 [1], Pythia 8.2 [2] [3] and Delphes 3.2 [4]. While MadGraph covered the simulation of matrix elements, Pythia was used to simulate the fragmentation and hadronization processes, and Delphes served as a detector simulator. The ISR jet production was controlled with Pythia, where the ISR emission was vetoed to one jet. Finally, in addition to the work done in each step of the chain, some scripts were written to integrate the programs. At this stage, *parallel* [5] was used to run simultaneous simulations.

B. Simulated channels

The values of the kinematic variables with which the histograms are filled are obtained from simulations of the SM channel:

$$p\ p \rightarrow t\ \tilde{t} + ISR\_jet \tag{16}$$

This sample document demonstrates proper use of REVTeX 4.1 (and LaTeX 2_ε) in mansucripts prepared for submission to APS journals. Further information can be found in the REVTeX 4.1 documentation included in the distribution or available at `http://authors.aps.org/revtex4/`.

When commands are referred to in this example file, they are always shown with their required arguments, using normal TeX format. In this format, `#1`, `#2`, etc. stand for required author-supplied arguments to commands. For example, in `\section{#1}` the `#1` stands for the title text of the author's section heading, and in `\title{#1}` the `#1` stands for the title text of the paper.

Line breaks in section headings at all levels can be introduced using \\. A blank input line tells TeX that the paragraph has ended. Note that top-level section headings are automatically uppercased. If a specific letter or word should appear in lowercase instead, you must escape it using `\lowercase{#1}` as in the word "via" above.

C. Second-level heading: Formatting

This file may be formatted in either the `preprint` or `reprint` style. `reprint` format mimics final journal output. Either format may be used for submission purposes. `letter` sized paper should be used when submitting to APS journals.

1. Wide text (A level-3 head)

The `widetext` environment will make the text the width of the full page, as on page **??**. (Note the use the `\pageref{#1}` command to refer to the page number.)

a. Note (Fourth-level head is run in) The width-changing commands only take effect in two-column formatting. There is no effect if text is in a single column.

D. Citations and References

A citation in text uses the command `\cite{#1}` or `\onlinecite{#1}` and refers to an entry in the bibliography. An entry in the bibliography is a reference to another document.

1. Citations

Because REVTeX uses the `natbib` package of Patrick Daly, the entire repertoire of commands in that package are available for your document; see the `natbib` documentation for further details. Please note that REVTeX requires version 8.31a or later of `natbib`.

a. Syntax The argument of `\cite` may be a single *key*, or may consist of a comma-separated list of keys. The citation *key* may contain letters, numbers, the dash (-) character, or the period (.) character. New with natbib 8.3 is an extension to the syntax that allows for a star (*) form and two optional arguments on the citation key itself. The syntax of the `\cite` command is thus (informally stated)

`\cite {` *key* `}`, or
`\cite {` *optarg+key* `}`, or
`\cite {` *optarg+key* `,` *optarg+key*... `}`,
where *optarg+key* signifies

key, or
`*`*key*, or
`[`*pre*`]`*key*, or
`[`*pre*`][`*post*`]`*key*, or even
`*[`*pre*`][`*post*`]`*key*.
where *pre* and *post* is whatever text you wish to place at the beginning and end, respectively, of the bibliographic reference (see Ref. [**?**] and the two under Ref. [**?**]). (Keep in mind that no automatic space or punctuation is applied.) It is highly recommended that you put the entire *pre* or *post* portion within its own set of braces, for example: `\cite { [ {`*text*`}]`*key*`}`. The extra set of braces will keep LaTeX out of trouble if your *text* contains the comma (,) character.

The star (*) modifier to the *key* signifies that the reference is to be merged with the previous reference into a single bibliographic entry, a common idiom in APS and AIP articles (see below, Ref. [**?**]). When references are merged in this way, they are separated by a semicolon instead of the period (full stop) that would otherwise appear.

b. Eliding repeated information When a reference is merged, some of its fields may be elided: for example, when the author matches that of the previous reference, it is omitted. If both author and journal match, both are omitted. If the journal matches, but the author does not, the journal is replaced by *ibid.*, as exemplified by Ref. [**?**]. These rules embody common editorial practice in APS and AIP journals and will only be in effect if the markup features of the APS and AIP BibTEX styles is employed.

c. The options of the cite command itself Please note that optional arguments to the *key* change the reference in the bibliography, not the citation in the body of the document. For the latter, use the optional arguments of the `\cite` command itself: `\cite *[`*pre-cite*`]` `[`*post-cite*`]{`*key-list*`}`.

[1] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, *et al.*, JHEP **1407**, 079 (2014), arXiv:1405.0301 [hep-ph].

[2] T. Sjstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, *et al.*, Comput.Phys.Commun. **191**, 159 (2015), arXiv:1410.3012 [hep-ph].

[3] T. Sjostrand, S. Mrenna, and P. Skands, JHEP **05**, 026 (2006), arXiv:hep-ph/0603175.

[4] J. de Favereau *et al.* (DELPHES 3), JHEP **1402**, 057 (2014), arXiv:1307.6346 [hep-ex].

[5] O. Tange, ;login: The USENIX Magazine **36**, 42 (2011).