DRAFT

Glen Cowan
RHUL Physics
3 February, 2009

Likelihood ratio for including b-tagging information

1 Introduction

Often a goodness-of-fit statistic such as Pearson's chi-squared variable is used to quantify the level of agreement between measured data and a given hypothesis. In an example of this considered here one tests whether a pair of jets is consistent with coming from a Higgs decay to $b\bar{b}$ for a specified Higgs mass $m_{\rm H}$ [1]. To test whether a jet pair is consistent with this decay, one can use the quantity

$$\chi^2 = \frac{(m - m_{\rm H})^2}{\sigma_m^2} \,, \qquad (1)$$

where m is the measured invariant mass of the jet pair and σ_m is the corresponding resolution. If m is Gaussian distributed and if the Higgs hypothesis is correct, then this statistic will follow a chi-square distribution for one degree of freedom.

Suppose that in addition to the mass, we also have tracking information that tells whether the jets contain particles with nonzero lifetime. This could be in the form of the p-value of the hypothesis that all of the particles in the jet originate from the primary event vertex. Such a variable will be uniformly distributed between zero and one if the hypothesis is true, and should peak near zero if the jet contains particles with nonzero lifetime, e.g., from the decay of b or c quarks. This is essentially what has been used in many b-tagging measurements at LEP (see, e.g., [2]). Here we will can this variable x.

For b-jets the distribution of x is peaked near its target value value of zero. One could try to add a term to the χ^2 statistic of the form

$$\frac{(x - \mu)^2}{\sigma_{x,b}^2} \qquad (2)$$

for each jet, where $\sigma_{x,b}$ is the standard deviation of x under the b-jet hypothesis and μ represents some central value about which x is distributed, e.g., the mean. But x is not Gaussian distributed about μ, and thus the term (2) will not follow a chi-squared distribution. Furthermore, a cut on χ^2 extended in this way does not represent the optimal separation between b-jets and those of other types.

In this note we describe a method for including the b-tagging information in a different way, namely, by using a likelihood ratio. The end result is a term that can be added to the χ^2 and which itself behaves like a chi-square goodness-of-fit statistic.

2 Distribution of the b-tagging variable

Suppose that we can estimate the distribution of x for different hypotheses, e.g., b-jets, c-jets and light-quark jets, e.g., using Monte Carlo or data control samples. Further let us assume that we can parameterize this distribution with some function $f(x;\theta)$, where θ is an adjustable parameter. As a simple example, we could consider a function such as

$$f(x;\theta) = (1+\theta)(1-x)^{\theta} \ . \tag{3}$$

For light-quark jets one would have $\theta = 0$, since there all of the tracks originate from the primary vertex. (It is assumed that K^0_S mesons and other long-lived strange hadrons have been excluded from the set of tracks considered.) Jets initiated by a b-quark could have, say, $\theta = 10$, and c-jets might have $\theta = 5$. These are purely hypothetical numbers used here for the sake of example. A plot of $f(x;\theta)$ for these values is shown in Fig. 1.

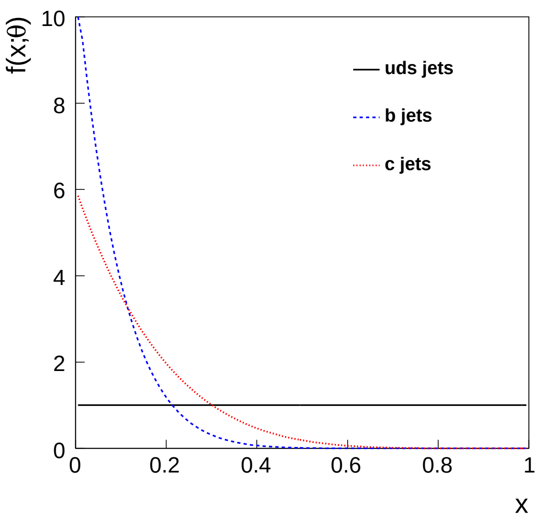

Figure 1: A possible parametric function to describe the distribution of the b-tagging variable x.

In practice the parametric function will not provide a perfect description of the x distribution, but for any given choice the best parameter values can be estimated for all of the hypotheses, e.g., b, c, and light (uds) quarks, which we can call θ_b, θ_c and θ_l. If the parametrization is imperfect then this will result in a less than optimal separation of the hypotheses, but in practice the degradation in performance may be small.

3 Likelihood ratio for the b-tagging variable

Suppose each jet in an event provides a measured x value. Consider first testing the hypothesis that all of the jets are all of the same flavour, i.e., that the x values all follow the pdf $f(x;\theta)$ for the same value of θ. The likelihood function for θ is

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) \ . \tag{4}$$

Using the pdf (3), the log-likelihood function is

$$\ln L(\theta) = \sum_{i=1}^{n} \left[\theta \ln x_i + \ln(\theta+1)\right] \ . \tag{5}$$

Setting the derivative of this with respect to θ equal to zero and solving gives the Maximum Likelihood (ML) estimator,

$$\hat{\theta} = -1 - \frac{n}{\sum_{i=1}^{n} \ln x_i} \; . \tag{6}$$

To quantify the level of agreement between an observed x and a hypothesized value of θ, we can construct the likelihood ratio

$$\lambda(\theta) = \frac{L(\theta)}{L(\hat{\theta})} \; , \tag{7}$$

The likelihood ratio lies in the range $0 \leq \lambda \leq 1$. If the data are in good agreement with the hypothesis, then $\hat{\theta}$ will be close to θ and thus λ is close to one. Small $\lambda(\theta)$ indicates poor agreement between the data and hypothesis. It is more convenient to work with the logarithmic variable

$$q_\theta = -2 \ln \lambda(\theta) \; , \tag{8}$$

where then $q_\theta \geq 0$ and increasing values indicate decreasing compatibility between data and hypothesis.

Under a set of regularity conditions and for a sufficiently large data sample, *Wilks' theorem* says that for a hypothesized value of θ, the pdf of the statistic $q_\theta = -2 \ln \lambda(\theta)$ approaches the chi-square pdf for one degree of freedom [3]. A proof and details of the regularity conditions can be found in standard texts such as [4].

Thus the quantity q_θ can simply be added to the original goodness-of-fit statistic (1). If the hypothesis considered is true, the extended statistic will follow a chi-square distribution for one additional degree of freedom.

3.1 Likelihood ratio for a single jet

Wilks' theorem says that $f(q_\theta|\theta)$ approaches a chi-square distribution in the large sample limit, i.e., when the number of observations of x is large. In many cases, however, one would like to carry out this procedure for a single jet (or perhaps for a pair of jets). In fact, we show here that q_θ is approximately chi-square distributed even for a single value of x, i.e., $n = 1$. In this case the ML estimator for θ is

$$\hat{\theta} = -1 - \frac{1}{\ln x} \; . \tag{9}$$

Single values of x were generated with Monte Carlo according to the pdf (3) using $\theta_b = 10$. Figure 2 shows the distribution of q_θ for test values $\theta = 10$ (the correct hypothesis) as well as $\theta = 5$ and 0.

A chi-square distribution for one degree of freedom is superimposed on Fig. 2. Ideally this should agree with the curve for $\theta = 10$ and it should do so when q_θ is based on a large number of observations of x. Here even with a single value of x the agreement is reasonably good.

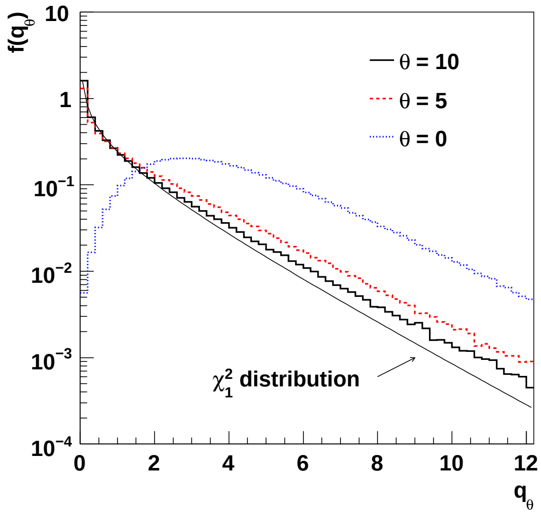

Figure 2: Distribution of the variable q_θ for $\theta = 0, 5, 10$ for data generated using $\theta_b = 10$. A chi-square distribution for one degree of freedom is also shown.

References

[1] Chris Parkinson and Veronique Boisvert, private communication.

[2] D. Buskulic et al. (The ALEPH Collaboration), Phys. Lett. B 313 (1993); see also David Brown and Markus Frank, *Tagging b hadrons using track impact parameters*, ALEPH 92-135, PHYSIC 92-124.

[3] S.S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, Ann. Math. Statist. **9** (1938) 60-2.

[4] A. Stuart, J.K. Ord, and S. Arnold, *Kendall's Advanced Theory of Statistics*, Vol. 2A: *Classical Inference and the Linear Model* 6th Ed., Oxford Univ. Press (1999), and earlier editions by Kendall and Stuart.