Nuclear Instruments and Methods in Physics Research A 830 (2016) 176–181

Contents lists available at ScienceDirect

Nuclear Instruments and Methods in Physics Research A

journal homepage: www.elsevier.com/locate/nima

Extrapolation technique pitfalls in asymmetry measurements at colliders

Katrina Colletti *, Ziqing Hong, David Toback, Jonathan S. Wilson

The George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, Texas A&M University, College Station, TX 77843-4242, United States

ARTICLE INFO

Article history:
Received 3 December 2015
Received in revised form
5 April 2016
Accepted 20 May 2016
Available online 25 May 2016

Keywords:
Asymmetry
Linear extrapolation
Monte Carlo
Collider experiments

ABSTRACT

Asymmetry measurements are common in collider experiments and can sensitively probe particle properties. Typically, data can only be measured in a finite region covered by the detector, so an extrapolation from the visible asymmetry to the inclusive asymmetry is necessary. Often a constant multiplicative factor is advantageous for the extrapolation and this factor can be readily determined using simulation methods. However, there is a potential, avoidable pitfall involved in the determination of this factor when the asymmetry in the simulated data sample is small. We find that to obtain a reliable estimate of the extrapolation factor, the number of simulated events required rises as the inverse square of the simulated asymmetry; this can mean that an unexpectedly large sample size is required when determining the extrapolation factor.

1. Introduction

Measurements of production asymmetries have a long history at colliders [1–21], so examination of the experimental techniques used to make them is important. Most measurements are performed by first measuring the asymmetry within a restricted geometric region—the region covered by the detector—and then extrapolating to the inclusive region. In some cases an extrapolation based on a constant multiplicative factor is advantageous, but a potential pitfall exists in estimating the multiplicative factor via simulations. Because this sort of technique is widely applicable to experimental measurements, we explore it in detail here and identify where and why this potential pitfall arises.

In general an asymmetry is defined with the partial cross sections, σ_1 and σ_2, over two complementary kinematic or geometric regions,

$$A \equiv \frac{\sigma_1 - \sigma_2}{\sigma_1 + \sigma_2}. \tag{1}$$

We can simplify our discussion by considering the regions defined by a single variable, x, while integrating over all other variables. In the case where x represents the pseudorapidity of a particle, which is directly related to the angle θ between an outgoing particle and the beam line, this produces a forward–backward asymmetry, for example, for use in top-quark-pair production at the Fermilab Tevatron [1–6]. We define $A^{\text{inclusive}}$ using

$$\sigma_1^{\text{inclusive}} = \int_0^{\infty} dx \frac{d\sigma}{dx}, \quad \text{and}$$
$$\sigma_2^{\text{inclusive}} = \int_{-\infty}^{0} dx \frac{d\sigma}{dx}. \tag{2}$$

However, when the entire range of x is not accessible due to kinematic constraints and/or the geometry of the detector, we can only measure

$$\sigma_1^{\text{visible}} = \int_0^{x^{\text{visible}}} dx \frac{d\sigma}{dx}, \quad \text{and}$$
$$\sigma_2^{\text{visible}} = \int_{-x^{\text{visible}}}^{0} dx \frac{d\sigma}{dx}, \tag{3}$$

which define the visible asymmetry, A^{visible}.

There are multiple ways to extrapolate from A^{visible} to $A^{\text{inclusive}}$. The two simplest methods for doing this are employing an additive correction factor ($C = A^{\text{inclusive}} - A^{\text{visible}}$) [22,23] or, a method that is commonly used, employing a multiplicative correction factor

$$R = \frac{A^{\text{visible}}}{A^{\text{inclusive}}}, \tag{4}$$

where each are typically estimated using Monte Carlo (MC) simulations [1,2,24]. Each is applicable in different physical scenarios. While more sophisticated correction methods can be, and in some cases must be, employed [4–21,25], the multiplicative correction method has been very successful for $t\bar{t}$ leptonic

* Corresponding author.
E-mail address: kcolletti1@tamu.edu (K. Colletti).

http://dx.doi.org/10.1016/j.nima.2016.05.082
0168-9002/

asymmetry measurements, as the correction factor appears not to vary significantly with the inclusive asymmetry [24]. In this paper, we explore a simple example in which this condition holds, but use it to identify a pitfall in the estimation of the correction factor and explore ways in which this pitfall may be avoided by future analyses.

For illustrative purposes, we consider a simplified model based on the measurement of the top leptonic forward–backward asymmetry at the Fermilab Tevatron [1–6]. It has been shown both that the differential cross section of leptons as a function of pseudorapidity can be well approximated as the sum of two Gaussian distributions with a common mean, and that the simple multiplicative extrapolation technique works in this case [24]. For the purposes of this study, we take the differential cross section $d\sigma/dx$ to be the simpler single-Gaussian distribution with unit width and a non-zero mean, μ. As shown in Appendix A there is an approximately linear relationship between the asymmetry and μ for small values of μ; we can refer to the behavior of μ and the asymmetry interchangeably. This simple model provides a foundation to understand the general behavior of multiplicative asymmetry extrapolation methods.

A potential pitfall occurs when estimating the correction factor in Eq. (4) using MC samples with small asymmetries. Under certain quantifiable conditions, simulations can produce values of R that are misleading and far from the correct value. To make the discussion concrete, we pick a visible region for our single Gaussian distribution of $-1.5 < x < 1.5$, which gives the visible and inclusive regions as shown in Fig. 1, with the dashed lines indicating the boundaries. Given this particular description, to an excellent degree of approximation we find $R = 0.7795 \pm 0.0005$, as shown in Appendix A. Since analyses typically have more complicated distributions and use MC methods to estimate R, we begin this study by using MC samples to determine the distribution of the multiplicative factor, and illustrate the pitfalls when the simulated $A^{\text{inclusive}}$ goes to zero. We then compare this result with a closed form statistical solution to gain a better understanding of why this pitfall arises.

2. Monte Carlo study

The most common method to determine the multiplicative correction factor is to simulate events according to a calculated differential cross section $\frac{d\sigma}{dx}$, and calculate the correction factor R from the simulated events. We mimic this procedure by generating sets of random numbers according to a simplified differential cross section that takes the form of a Gaussian function with unit width and a mean μ. Each random number represents an event, each set of random numbers is a pseudo-experiment (PE), and the number of events in each PE is denoted by N. From each PE, we can measure both A^{visible} and $A^{\text{inclusive}}$, and therefore R. Distributions of these three values can then be generated with an ensemble of PEs; the number of PEs used to generate these distributions is denoted by N_{PE}. For example, in Fig. 1, we show two examples of differential cross sections (a single PE) with $N = 10^6$ and both $\mu = 0.0$ and $\mu = 0.5$. In Fig. 2, we show the distributions of A^{visible}, $A^{\text{inclusive}}$, and R for $N_{\text{PE}} = 10^6$, each with $N = 10^6$ and $\mu = 0.1$. This value of μ is chosen as it corresponds to $A^{\text{inclusive}} \approx 8\%$, which is a value we typically see in $t\bar{t}$ asymmetry measurements at the Tevatron [1–6].

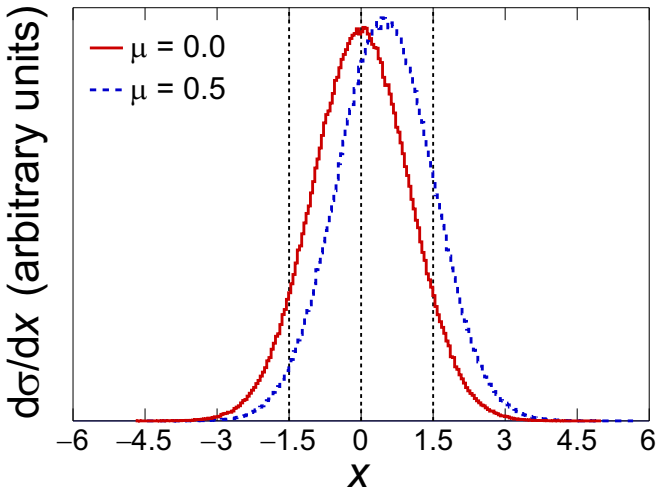

Fig. 1. Two Gaussian distributions with unit width, with $\mu = 0.0$ and $\mu = 0.5$. The dashed lines at -1.5 and 1.5 indicate the boundaries for the visible region we consider here.

Since the simulation of a practical differential cross section is usually computationally expensive, the common practice is to simulate one PE with a modest N, usually on the order of 10^6, and calculate R from it. In this analysis, the distribution of R from an ensemble of PEs reveals the quality of the estimation of R from a single PE. We note that in Fig. 2(c) the variation in R is small, with a width less than 1% of its mean value. With the simplified single-Gaussian differential cross section and the visible region specified above, R=0.7798 which is consistent with the calculation in Appendix A.

We next study the quality of the estimation of R as we vary the two factors, μ and N, which have significant impact on potential measurements: we examine what happens both in the limit of small simulation sample size and as $\mu \to 0$ (or equivalently, as the asymmetry approaches zero). Specifically, we aim to understand whether the estimation of R is correct and what the uncertainty on that estimation is, the sample size needed to obtain a small uncertainty, and whether the value of R is constant for all values of μ when it is measured with a large sample size.

For $\mu = 0.1$, R is well determined even with a fairly small value of N. Fig. 3 shows distributions of R for $N_{\text{PE}} = 10^6$ with $N = 10^5$ and $N = 10^3$. As N decreases, the R distribution widens and becomes much less Gaussian, and the peak of the distribution shifts. Thus, estimating the value of R from a single PE (as is typically done in realistic scenarios with more complicated differential cross sections) quickly leads to incorrect results. That is, there is a minimum allowable N, above which we can be confident in the estimation of R, and below which the estimation of R is no longer reliable, hence introducing a significant and asymmetric systematic uncertainty to the inclusive asymmetry measurement. This issue becomes even more pronounced as μ, and thus the asymmetry, approaches 0.

As we note in the next section, the statistics behind this effect has been studied in great detail in the literature, and we see that the R distribution begins to approximate a Cauchy distribution [26]. To explain this, we note that the Cauchy distribution is the distribution of the ratio of two Gaussian random variables when the mean of the denominator is zero. When the mean of the Gaussian in the denominator is far enough away from zero, the distribution is Gaussian, and in the limit that it approaches zero, the distribution approaches the Cauchy distribution. The usual measurements of mean and standard deviation are not expected to give accurate and reliable results; indeed, for a true Cauchy distribution these two values are not defined.

To determine how many events we need to be able to make a reliable estimation of R, we define the fraction of PEs with $R < 0.5$:

$$f = \frac{N_{\text{PE}}(R < 0.5)}{N_{\text{PE}}(\text{total})}. \quad (5)$$