Extrapolation Techniques for Asymmetry Measurements

Katrina Colletti

Z. Hong, D. Toback, J.S. Wilson
Texas A&M University
New Perspectives Meeting, Fermilab

kcolletti1@tamu.edu

June 8, 2015
Overview

1. Background, Motivation, and Goals

2. Study 1: Monte Carlo Simulation

3. Study 2: Closed Form Statistical Solution

4. Study 3: Closed Form Numerical Solution

5. Conclusions
Common in particle physics to measure asymmetries – in particular in collider experiments.

Often data can only be measured for a finite portion of the detector, must extrapolate to the total asymmetry.

\[
A^\text{total} = \frac{(C + D) - (A + B)}{A + B + C + D}
\]

\[
A^\text{finite} = \frac{C - B}{B + C}
\]

Can we use a simple constant multiplicative factor \(A^\text{total} = R \cdot A^\text{finite}\)?

If so, how much statistics needed to get reliable results, especially in the limit of small asymmetries?

Classic Example: forward-backward asymmetry (\(A_{FB}\)) measured in collider detectors:

- Positive Beam Direction
- Outgoing particle momentum
- \(\theta_{CM}\)
- Events = 1000000

Gaussian Distribution, Mean = 0
We start with a single Gaussian with a mean of $\mu$ as a good working model to build a foundation and give good insights into more complicated distribution models.

Examples from collider physics have shown that this approximation sometimes works.

It is not obvious if a linear extrapolation technique should work.

Since we typically use MC methods to estimate such values, we need to understand whether we can confidently use a constant $R$ to linearly extrapolate, and understand the amount of statistics needed to get a reasonable measurement of it.
Study 1: Monte Carlo Simulation

- In our simple Gaussian model, $A$ is linearly proportional to $\mu$ (the mean of the distribution)
- Example: $\mu = 0.1$ corresponds to $A^{\text{total}} \approx 8\%$ which is what we typically see in forward-backward asymmetry top quark measurements at the Tevatron
- Run many MC pseudo-experiments each with a large number of events, get distributions for $A^{\text{total}}$, $A^{\text{finite}}$, and $R$:
Study 1: Monte Carlo Simulation

- With enough statistics (i.e. large $N$), measurements of $R$ are very accurate.
- As $N$ decreases, measurement of $R$ becomes unreliable, and can no longer correctly reproduce $A^{\text{total}}$ from $A^{\text{finite}}$.
- This is observed for all values of $\mu$.
With this understanding, we now aim to quantify this behavior to properly understand how many MC events in the original distribution, $N$, are needed to give reliable measurements of $R$.

We define $f$ as the fraction of pseudo-experiments with $R < 0.5$ (very far from expected value).
Study 1: Monte Carlo Simulation

- Want $f \approx 0$, define a threshold value and observe the relationship between the number of events needed for reliable measurements and $\mu$
- $N$ falls as $\frac{1}{\mu^2}$

- Measurements of $R$ for all values of $\mu$ with enough statistics give the same value
- Conclusion is that $R$ is indeed constant for all $\mu$ for this simple Gaussian model, and a huge amount of MC statistics are needed to accurately measure the actual value for small $\mu$ (or equivalently small $A$)
Study 2: Closed Form Statistical Solution

- Let’s take a closer look at *why* the MC methods break down

- Require $A_{total}^\text{total}$ (denominator of $R$) to be greater than *at least* $1\sigma$ away from 0 – to avoid the potential divide by 0 problem (math jargon: this is where the distribution transitions to a Cauchy regime)
The statistical question becomes: how many events, \( N \), are required for the mean of \( A_{FB}^{\text{total}} \) to be some number \(( k \cdot \sigma )\) away from 0, thus giving reliable measurements

\[
\sigma_{A_{FB}^{\text{total}}} = \frac{A_{\text{total}}}{k}
\]

Using statistics (see backup slides), we are able to find \( N \) as a function of \( \mu \) for our single Gaussian model:

\[
N = 2k^2 \cdot \frac{\left(1 + \text{erf} \left( \frac{\mu}{\sqrt{2}} \right) \right)}{\text{erf} \left( \frac{\mu}{\sqrt{2}} \right)^2}
\]

Some limiting cases:

- As \( \mu \to 0 \), \( N \to \infty \)
- Using the approximation \( \text{erf} \left( \frac{\mu}{\sqrt{2}} \right) \approx \sqrt{\frac{2}{\pi}} \mu \) for small \( \mu \), we find that \( N \propto \frac{1}{\mu^2} \) which is precisely what we just saw from our MC study
Closed form solution: blue (for $k = 2$)

MC data: red

Excellent agreement!
Study 3: Closed Form Numerical Solution

- We calculate $R$ as a function of $\mu$ using Mathematica
- Set $\sigma = 1.0$
- Plot $R$ in the limit $\mu \to 0$
- For large values of $\mu$, $R$ only rises by 0.04% relative to $\mu = 0$

\[
A_{total} = \frac{(C + D) - (A + B)}{A + B + C + D}
\]

\[
A_{finite} = \frac{C - B}{B + C}
\]

\[
R = \frac{A_{finite}}{A_{total}}
\]
Conclusions

- We have used three methods to study the linear extrapolation of $A_{\text{finite}}$ to an inclusive $A_{\text{total}}$.
- While we have only studied the simple Gaussian model, we observed that a linear extrapolation can be used, and while MC methods work reliably (even for small $A$) they can require much more significant statistics than expected.
- Our results have the potential to be applied for many different asymmetry measurements in collider experiments, and have already been useful at the Tevatron for the $t\bar{t}$ forward-backward asymmetry.
Thank You For Listening!
Any Questions?
We need enough statistics such that $A_{FB}^{total}$, the denominator of $R$, is more than 1 sigma away from 0 (we will set it to be $k$, where $k$ will be determined later). In other words, we want to know how many events it takes in a pseudo-experiment to ensure the mean of the full asymmetry will be $k$ standard-deviations away from zero. To do this we start with the equation

$$\sigma_{A_{FB}^{total}} = \frac{A_{FB}^{total}}{k}$$

(1)

where $\sigma_{A_{FB}^{total}}$ is the variation (or uncertainty) of the measured value of $A_{FB}^{total}$. We will find both $\sigma_{A_{FB}^{total}}$ and $A_{FB}^{total}$ as functions of $N$ and $\mu$ and substitute them into Eq. 1 to get the functional relation between $N$ and $\mu$ for “good statistics”. 
We begin with our definition of asymmetry,

\[ A_{FB}^{total} = \frac{N_+ - N_-}{N_+ + N_-} \]  

(2)

where \( N_+ = C + D \) and \( N_- = A + B \) as on Slide 2. Next we define \( N = N_+ + N_- \) as the total number of events in the original Gaussian distribution, and rewrite this as:

\[ A_{FB}^{total} = \frac{2N_+ - N}{N}. \]  

(3)

We note that since our distributions are Gaussian, we can write \( N_+ \) in terms of \( N \) and \( \mu \), with the relation given by

\[ N_+ = \frac{N}{\sqrt{2\pi}} \int_{0}^{\infty} dx \ e^{-(x-\mu)^2/2} \]

\[ = \frac{N}{2} \left( \text{erf} \left( \frac{\mu}{\sqrt{2}} \right) + 1 \right) \]  

(4)
Plugging this in to Eq. 3 and reducing, we get

\[
A_{FB}^{total} = \frac{2^{N/2} \left( \text{erf} \left( \frac{\mu}{\sqrt{2}} \right) + 1 \right)}{\mathcal{N}} \left( \text{erf} \left( \frac{\mu}{\sqrt{2}} \right) + 1 \right) - \mathcal{N}
\]

\[
= \text{erf} \left( \frac{\mu}{\sqrt{2}} \right)
\]

We next find \( \sigma_{A_{FB}^{total}} \) by beginning with the definition given in Bevington (92) applied to our problem,

\[
\sigma_{A_{FB}^{total}} = \left( \frac{\partial A_{FB}^{total}}{\partial N_+} \right) \sigma_{N_+} + \left( \frac{\partial A_{FB}^{total}}{\partial N} \right) \sigma_{N}.
\]

Taking a simple derivative of \( A_{FB}^{total} \) from Eq. 3 gives us

\[
\left( \frac{\partial A_{FB}^{total}}{\partial N_+} \right) = \frac{2}{N}
\]
To be consistent with the previous study, we fix $N$ and allow $N_+$ to vary. This means that $\sigma_N = 0$, and from simple statistics

$$\sigma_{N_+} = \sqrt{N_+}$$

(8)

Plugging Eqs. 7 and 8 into Eq. 6, we get

$$\sigma_{A_{FB}}^{\text{total}} = \frac{2}{N} \cdot \sqrt{N_+}.$$  

(9)

Plugging Eq. 4 into this, we get

$$\sigma_{A_{FB}}^{\text{total}} = \frac{2}{N} \cdot \sqrt{\frac{N}{2}} \left( \text{erf} \left( \frac{\mu}{\sqrt{2}} \right) + 1 \right)$$

$$= \sqrt{\frac{2}{N}} \cdot \sqrt{\left( 1 + \text{erf} \left( \frac{\mu}{\sqrt{2}} \right) \right)}.$$  

(10)
Finally, plugging Eqs. 5 and 10 back into Eq. 1 gives us

$$\sqrt{\frac{2}{N}} \cdot \sqrt{\left(1 + \text{erf}\left(\frac{\mu}{\sqrt{2}}\right)\right)} = \frac{\text{erf}\left(\frac{\mu}{\sqrt{2}}\right)}{k},$$  \hspace{1cm} (11)

and solving for $N$, we get

$$N = \frac{2k^2 \left(1 + \text{erf}\left(\frac{\mu}{\sqrt{2}}\right)\right)}{\text{erf}^2\left(\frac{\mu}{\sqrt{2}}\right)}.$$  \hspace{1cm} (12)

This is, as we set out to solve for, the number of events it takes per pseudo-experiment to ensure the mean of the full asymmetry will be $k$ standard-deviations away from zero, and thus give good statistics. Discussion of the implication of this result is included in the main slides.
\[ A_{FB}^{total} = \frac{1}{\sqrt{2\pi}\sigma} \int_0^\infty dx \left[ \exp\left( -\frac{(x-\mu)^2}{2\sigma^2} \right) - \exp\left( -\frac{(-x-\mu)^2}{2\sigma^2} \right) \right] \]
\[ A_{FB}^{finite} = \frac{1}{\sqrt{2\pi}\sigma} \int_0^{1.5} dx \left[ \exp\left( -\frac{(x-\mu)^2}{2\sigma^2} \right) - \exp\left( -\frac{(-x-\mu)^2}{2\sigma^2} \right) \right] \]
\[ R = \frac{A_{FB}^{finite}}{A_{FB}^{total}} \]