CDF/MEMO/ONLINE/GROUP/10883
Version 0.0
July 31, 2012

Statistical Inference by Orthogonal Functions: Applications to High-Energy Physics

N. F. Kazmi [1]and K. J. Knoepfel[2]

Fermi National Accelerator Laboratory

1 Introduction

2 Expansions in terms of moments

Any function $f(x)$ can be written as the series expansion

$$f(x) = \sum_{n=0}^{\infty} c_n g_n(x) \tag{1}$$

where the set $\{g_n(X)\}$ must be complete in order for the equality to hold. A very common assignment is $g_n(x) = (x - x_0)^n$, in which case, the coefficients are determined by taking successive derivatives of $f(x)$, leading to the well-known Taylor expansion:

$$f(x) = \sum_{n=0}^{\infty} \frac{1}{n!}(x - x_0)^n \frac{d^n}{dx^n} f(x_0) \tag{2}$$

We now consider the situation where $f(x)$ represents a p.d.f. $f(X)$, where X is a random variable, subject to the normalization constraint:

$$\int_a^b f(X)\; dX = 1\ . \tag{3}$$

For a p.d.f. with support in the range $X \in \mathbb{R}$, the integration is over all $\mathbb{R}$, in which case the individual terms in the Taylor expansion are divergent.

In general, a truncated Taylor expansion of a probability density function $f(X)$ cannot be used to approximate f as integrating the series term-by-term yields divergent results. We shall define the characteristic function (c.f.) as the Fourier transform of a p.d.f.:

$$\phi_X(u) = \int e^{iXu} f(X)\; dX \tag{4}$$

[1]nkazmi@fnal.gov
[2]knoepfel@fnal.gov

The c.f. always exists for any p.d.f. as it is bounded by the Cauchy-Jordan lemma:

$$\phi_X(u) \leq \int e^{iXu} dX \int f(X)\ dX = 1 \tag{5}$$

One can Taylor-expand the exponential such that the c.f. can be written as an infinite series, given by:

$$\phi_X(u) = \sum_{n=0}^{\infty} \frac{i^n}{n!} \mu'_n u^n\ , \tag{6}$$

where μ'_n is the n^{th}-order algebraic moment. Note that this expansion is only defined for distributions where the moments μ'_n exist. If one wishes to revert back to the p.d.f., the inverse Fourier transform can be taken such that:

$$f(X) = \mathcal{F}^{-1}\left[\phi_X(u)\right] \tag{7}$$

$$= \frac{1}{2\pi} \int e^{-iXu} \phi_X(u)\ du \tag{8}$$

Using the expansion in Eq. (6), the p.d.f. can be written as

$$f(X) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \mu'_n \delta^{(n)}(X) \tag{9}$$

where the infinite numbers of singularities implies that the p.d.f. $f(X)$ can only be understood in a distributional sense — i.e. by calculating moments of the distribution. Although the expression in Eq. (9) is now written in terms of the moments of the distribution, the presence of the delta functions makes usage of Eq. (9) impractical.

3 Gram-Charlier expansions

Let us take the following example, where we assume that a function can be expanded based on perturbations about a generating function $g_n(X)$:

$$f(X) = \sum_{n=0}^{\infty} c_n \frac{d^n}{dX^n} g_n(X) \tag{10}$$

The generating function has the property that:

$$\frac{d^n}{dX^n} g_n(X) = P_n(X) h(X) \tag{11}$$

where $P_n(X)$ is an n^{th}-degree orthogonal polynomial, and $h(X)$ is a residual function that remains and is independent of the order n. Equation (12) can now be rewritten as:

$$f(X) = \sum_{n=0}^{\infty} c_n P_n(X) h(X) \tag{12}$$

To determine the value of the c_n coefficients, both sides are multiplied $P_m(X)w(X)/h(X)$ and integrated over the relevant range for X:

$$c_n = \frac{1}{U_n} \int \frac{w(X)}{h(X)} P_n(X) f(X)\, dX \tag{13}$$

where U_n is the normalization factor from the orthogonality relationship:

$$\int w(X) P_n(X) P_m(X)\, dX = U_n \delta_{nm} \tag{14}$$

where $w(X)$ is the weighting function wrt. the chosen orthogonal polynomial, and δ_{nm} is the Kronecker delta function.

We then express the polynomials in terms of the power series:

$$P_n(X) = \sum_{m=0}^{n} a_m X^m \tag{15}$$

resulting in the expression:

$$c_n = \frac{1}{U_n} \sum_{m=0}^{n} a_m E\left[X^m \frac{w(X)}{h(X)}\right] \tag{16}$$

where the expectation value is

$$E\left[X^m \frac{w(X)}{h(X)}\right] = \int X^m \frac{w(X)}{h(X)} f(X)\, dX. \tag{17}$$

Note that in the event that $w(X) = h(X)$, the algebraic moments $E[X^m] = \mu'_m$ are obtained, and the expression for the p.d.f. now reads:

$$f(X) = \sum_{n=0}^{\infty} \frac{1}{U_n} \left[\sum_{m=0}^{n} a_m \mu'_m\right] P_n(X) h(X) \tag{18}$$

where the a_m coefficients are the same ones that appear in the definitions of the orthogonal polynomials $P_n(X)$. The double series can then be rearranged so that n refers to the order of the moment, and not the polynomial:

$$f(X) = \sum_{n=0}^{\infty} \left[\frac{h(X)}{U_n} \sum_{m=0}^{n} b_{mn} P_m(X)\right] \mu'_n \tag{19}$$

where the b_{mn} coefficients in general also depend on n. The next section lists the relevant coefficients for various orthogonal polynomials.

4 Fits to spectra

Fits are provided at the end of this section.

4.1 Gram-Charlier Type A Expansion

There were two distinct methods used in this analysis. The first is was the Gram-Charlier Type A Expansion.

$$\int_{-\infty}^{\infty} He_m(X)\, He_n(X)\, e^{-\frac{X^2}{2}}\, dX = \sqrt{2\pi}\, n!\, \delta_{nm} \tag{20}$$

The formula to solve for each polynomial:

$$H_n(X) = n! \sum_{m=0}^{n/2} \frac{(-1)^m}{m!\,(n-2m)!}\, 2^{-m}\, X^{n-2m} \tag{21}$$

This formula loops through a set of integers even though this type of polynomial should exist over all real numbers. The initial attempted method was using the Gram-Charlier definition. For the following explanations Order refers to the Order of the Hermite Polynomial and the Moment.

Using the Gram-Charlier Expansion [1] the probability function was able to fit the JetEnergy1 Distribution, Figure(1). As the Order increases the function converged to the distribution. Order 5, Figure(4) at low Orders the distribution is not well defined. The fluctuations in the probability. function come from its Moments. The Moments wobble, creating the fluctuations.

The Moments in this and following plots show Moments 0-5. At Order 5, Moment 0 plays the largest roles, Figure(5)), this is why at lower Moments the probability curve is almost Gaussian in nature.

As the Order increases, (Figure(6), now at Order 10) the probability distribution becomes more accurate. There are still fluctuation at the tail end of the function. The larger Moments go into affect as well. At Order 10, Figure(7), the Gaussian is not as important and instead the higher Moments are able to shift the function so that it will converge.

The fluctuations at the tail end of the function fade away as Order is set to 20. Figure(8) is of Order 20. The wobbles at the tail end of the function are still apparent at higher Orders but they are not as affective. Figure(9), shows that the higher Moments taper off with a fewer wobbles.

$$F(x) \approx \frac{1}{\sqrt{2\pi}\sigma} exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]\left[1 + \frac{k_3}{3!\sigma^3} H_3\left(\frac{x-\mu}{\sigma}\right) + \frac{k_4}{4!\sigma^4} H_4\left(\frac{x-\mu}{\sigma}\right)\right] \tag{22}$$

4.2 Gauss-Hermite Expansion

The second method is the the Gauss-Hermite Expansion.

$$\int_{-\infty}^{\infty} H_m(X)\, H_n(X)\, e^{-X^2}\, dX = \sqrt{\pi}\, 2^n\, n!\, \delta_{nm} \tag{23}$$

The formula to solve for each polynomial:

$$H_n(X) = n! \sum_{m=0}^{n/2} \frac{(-1)^m}{m!\,(n-2m)!}\,(2X)^{n-2m} \tag{24}$$

Traditionally Hermite Polynomials exist over all real numbers. However in this situation, the polynomials were limited to a specific region.
The Gram-Charlier Expansion works relatively well but it converges at much higher Orders, so it takes longer to calculate. To make the function converge faster a new method was attempted. The next step was to move to the Gauss-Hermite Expansion definition [1]. These polynomials converged at lower Orders.

At Order 5, Figure(10), the Gauss-Hermite Expansion converges to the distribution, however the χ^2 function is still high. Unlike the Gram-Charlier Expansion, Order 5, the Gauss-Hermite Expansion function does not have large fluctuations at the tail of the distribution. The Moments for the Gauss-Hermite Expansion have the same trend as the Moments for the Gram-Charlier Expansions. Meaning that for lower Orders the Gaussian plays a larger role and at higher Orders the distribution becomes more asymmetric to fit the jetenergy distribution. Moment at Order 5, Figure(11), shows that Moment 0 has the largest affect.

At Order 10, Figure(12), the function is relatively similar to the jet energy distribution. The fluctuations have nearly gone away as the function converges. Order 10, Figure(13) shows that the Moments of the distribution wobble at wildly. The extremes highs and lows cancel out creating the nearly smooth probability function.

The precision of the approximation becomes more accurate as the Order increases At Order 20, Figure(14) converges with the distribution very well. Simply comparing Order 20 to Order 5 and Order 10 (or any of the probabilists distributions) shows this as well. Order 20, Figure(15), the distribution of the Moments follows the same trend.

4.3 Differences in Methods

It is important to notice that the calculations between Gram-Charlier and Gauss-Hermite change. To go from using the Gauss-Hermite definition to the Gram-Charlier definition, in Eq.(24) the values of 2X must be replaced with $(\sqrt{2})X$ and the formula must be multiplied by $2^{-\frac{n}{2}}$. The final equation will be Eq.(21).

The intended use for either method was to insert them in the following equation.

$$p(X) = \left[1 + \frac{1}{3!}\mu_3 H_3(X) + \frac{1}{4!}(\mu_4 - 3)H_4(X) + ...\right]\Phi(X) \tag{25}$$

This probability function is a series that involves increasing orders of Moments and Hermite Polynomials.

$$c_n = \frac{1}{n!}\int_{-\infty}^{\infty} p(X)\, H_n(X)\, dX \tag{26}$$

$$c_n = \frac{1}{n!}\left[\mu_n - \frac{n^{[2]}}{2 \times 1!}\mu_{n-2} + \frac{n^{[4]}}{2^2 \times 2!}\mu_{n-4} - \frac{n^{[6]}}{2^3 \times 3!}\mu_{n-6} + ...\right] \tag{27}$$

Cumulants were neccessary for the calculation of this process. Though, the cumulants had to be adjusted. The formula for cumulants is in Eq.(26). When it is expanded it follows the series Eq.(27) where $n^{[m]} = \frac{n!}{(n-m)!}$. There was an issue with strictly these equations. In the Gram-Charlier calculation, the correct solution was to include moments of smaller orders until the order on each coefficient was the same. For the Gauss-Hermite it was to include a mu_2^z until the order of the moments multiplied by each coefficient was the same.

The Gram-Charlier and the Gauss-Hermite Expansions were applied to a specific distributions of JetEnergy1,Figure(1). Gram-Charlier was also applied to JetEnergy2, Figure(2). The values in these distributions were merged and calculated to create a solution for JetEnergyHT Figure(3).

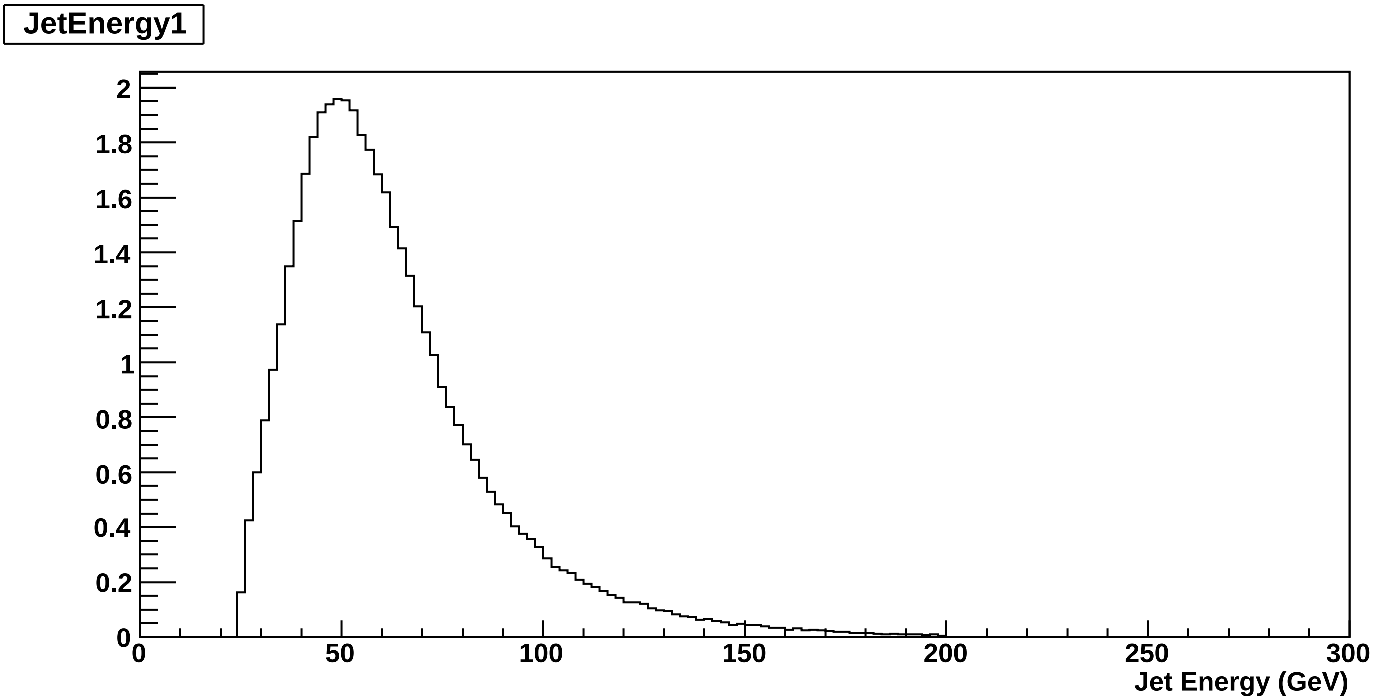

Figure 1: Distribution of JetEnergy1

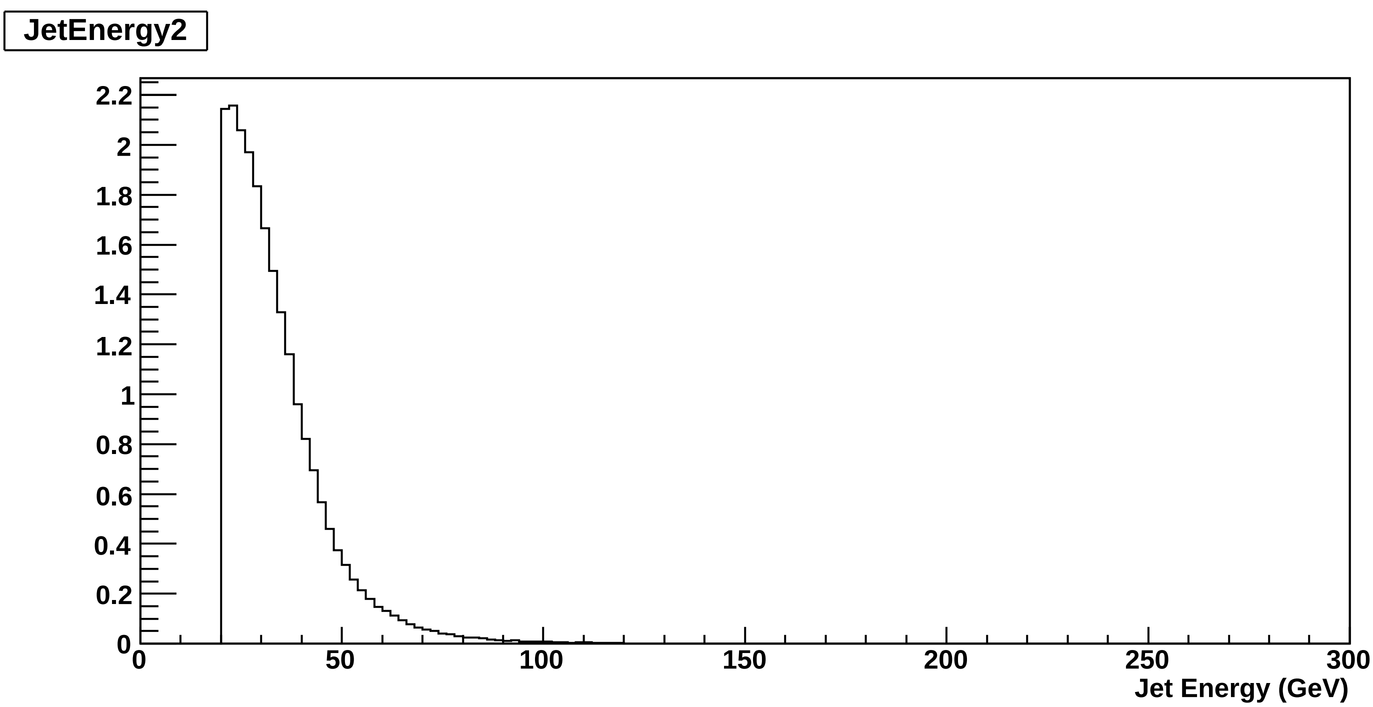

Figure 2: Distribution of JetEnergy2

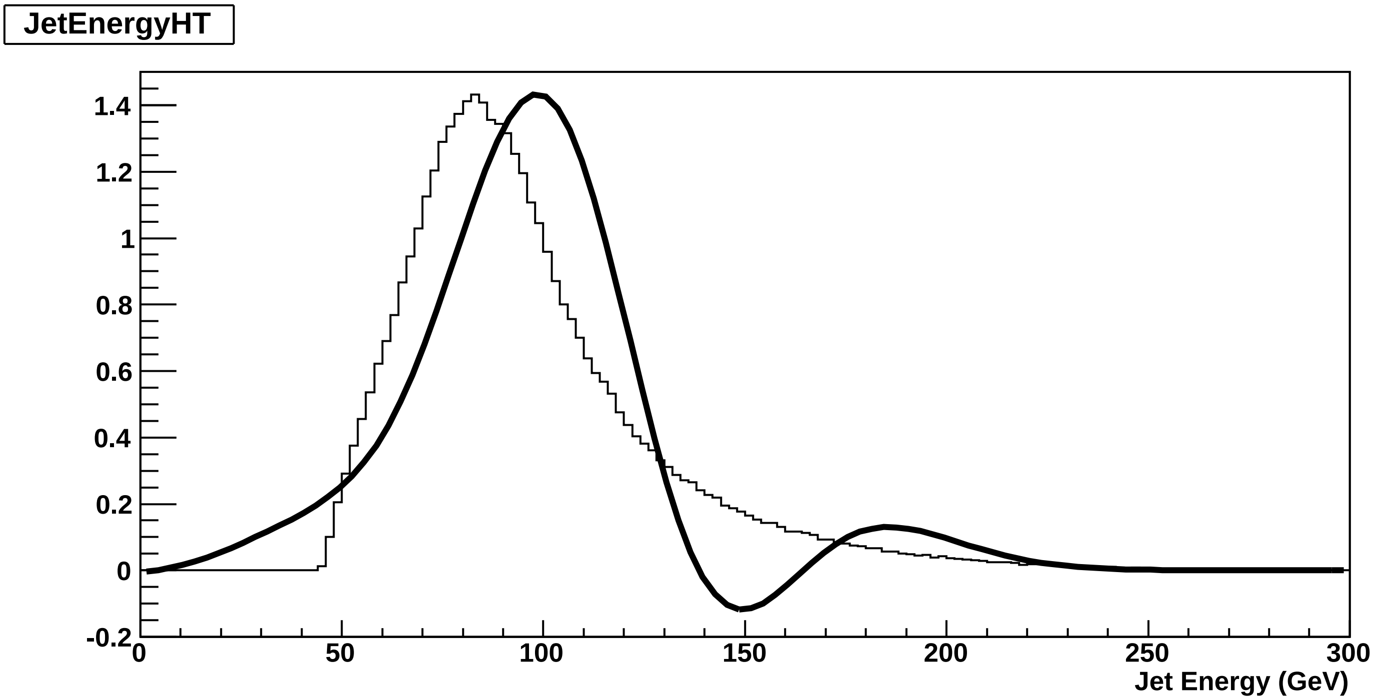

Figure 3: Calculating JetEnergyHT from JetEnergy1 & JetEnergy2.

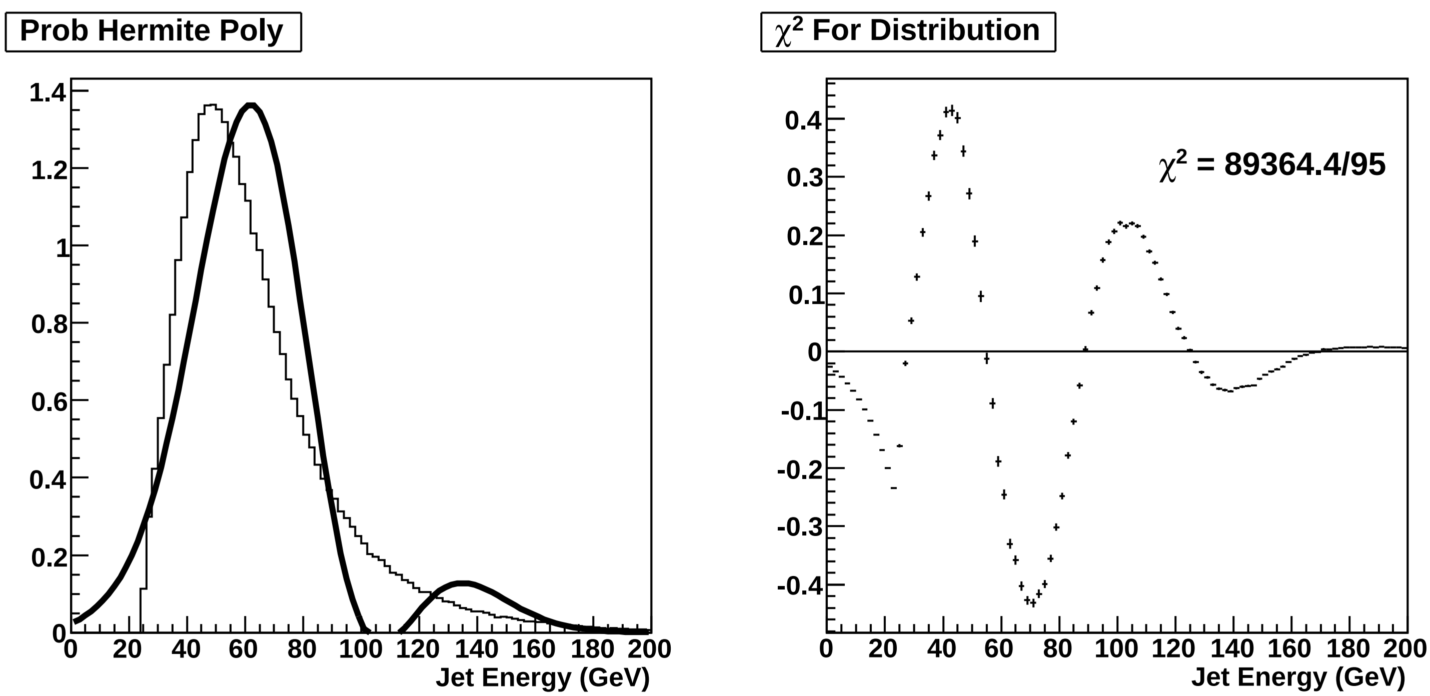

Figure 4: Calculating Probability using Moments and Prob Hermite Poly. Order 5

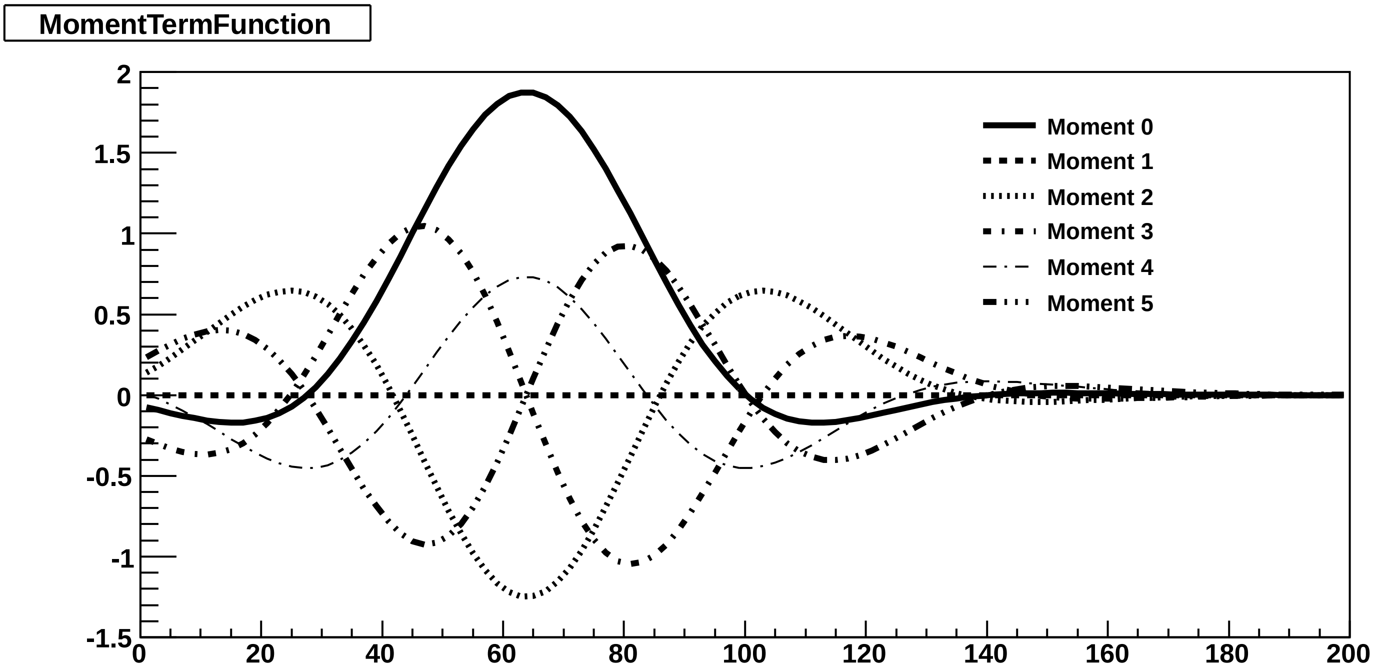

Figure 5: Calculating Probability using Moments and Prob Hermite Poly. Order 5

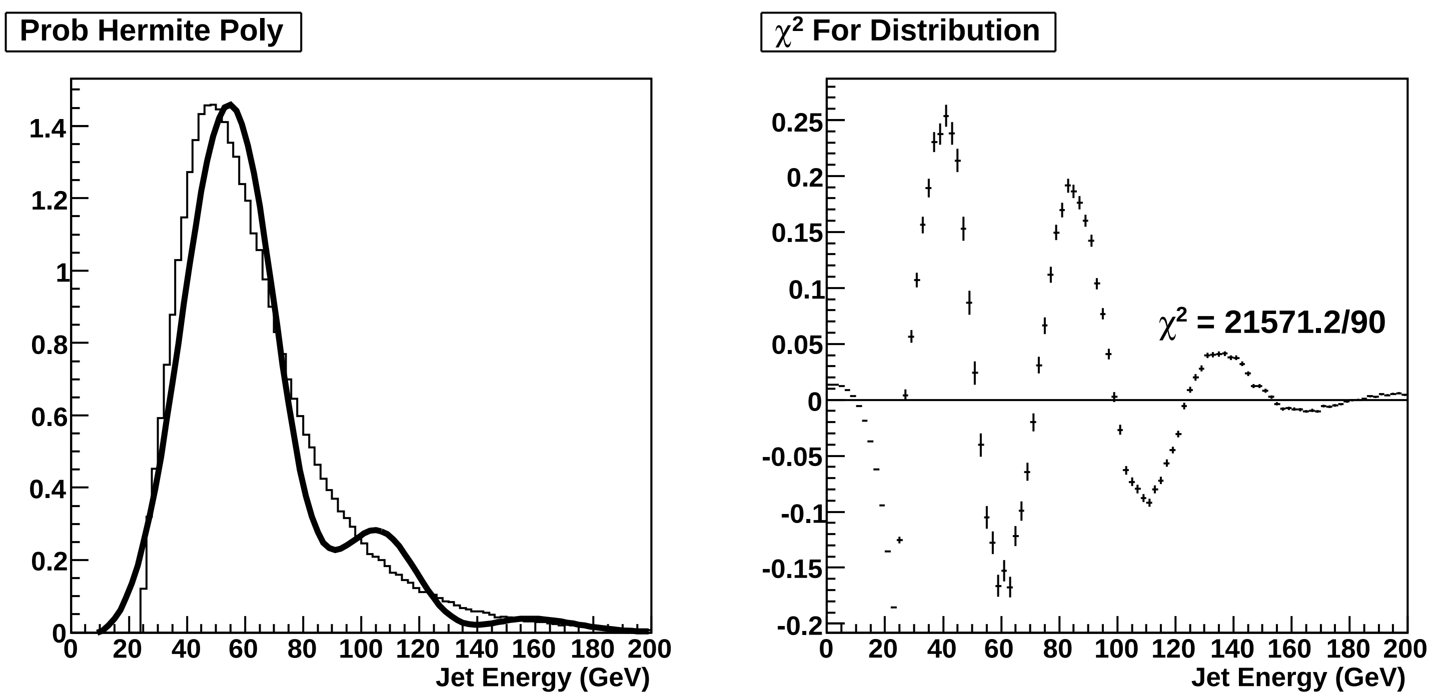

Figure 6: Calculating Probability using Moments and Prob Hermite Poly. Order 10

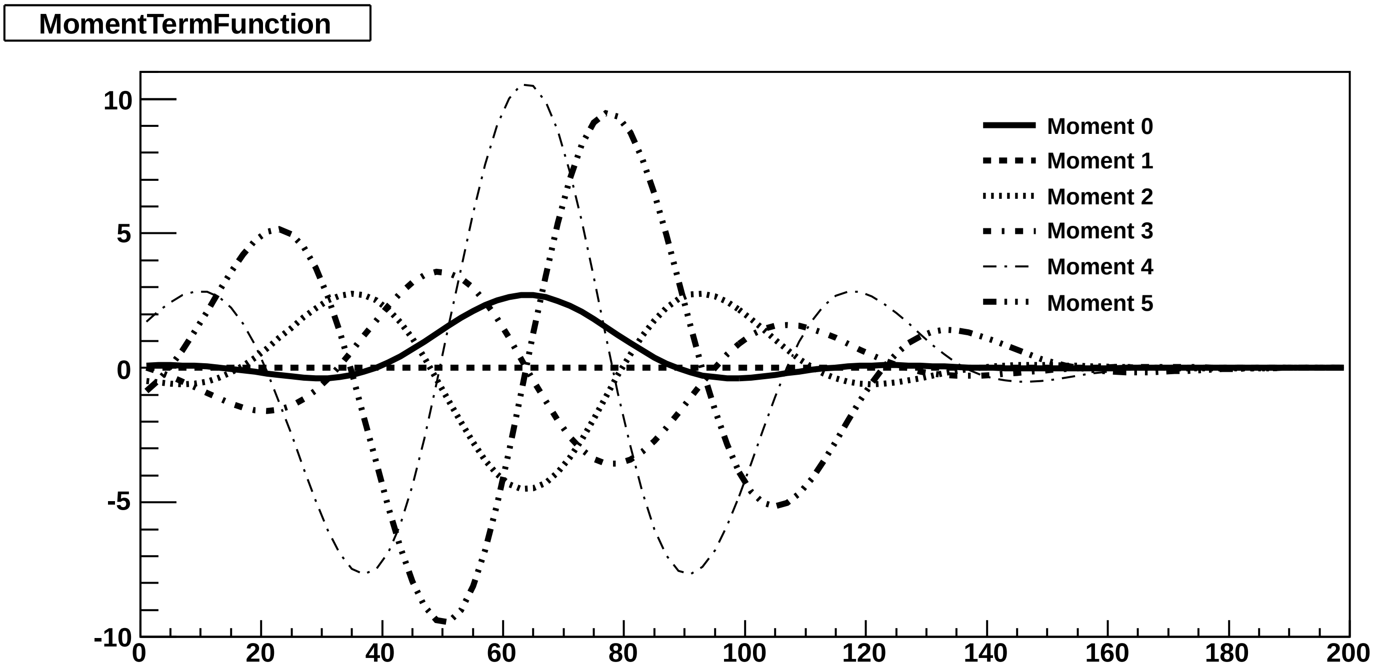

Figure 7: Calculating Probability using Moments and Prob Hermite Poly. Order 10

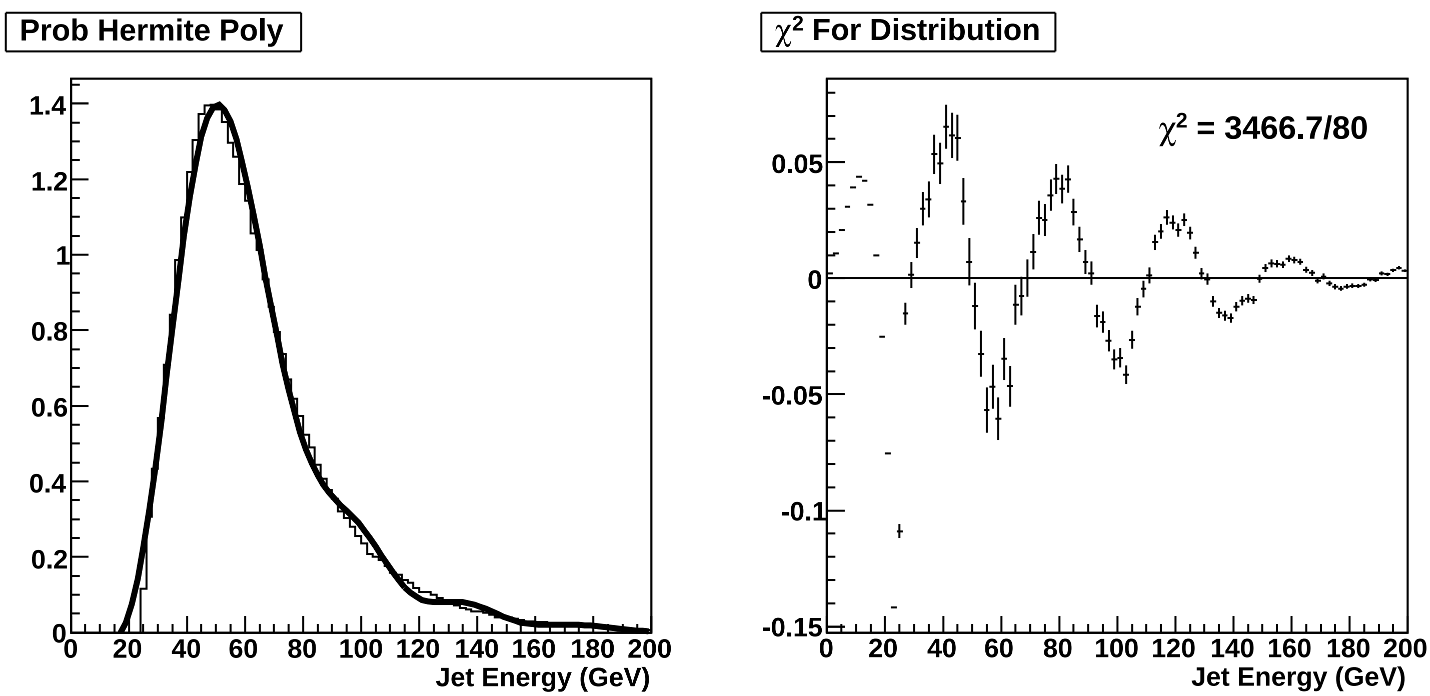

Figure 8: Calculating Probability using Moments and Prob Hermite Poly. Order 20

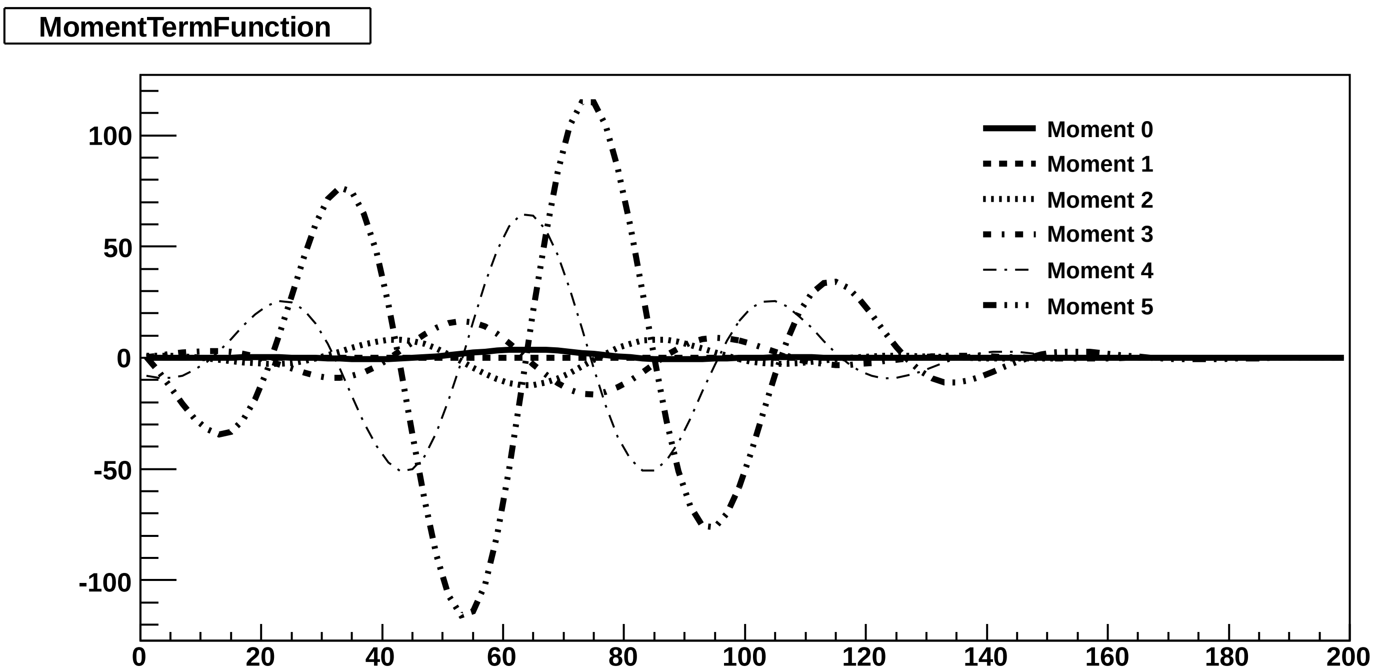

Figure 9: Calculating Probability using Moments and Prob Hermite Poly. Order 20

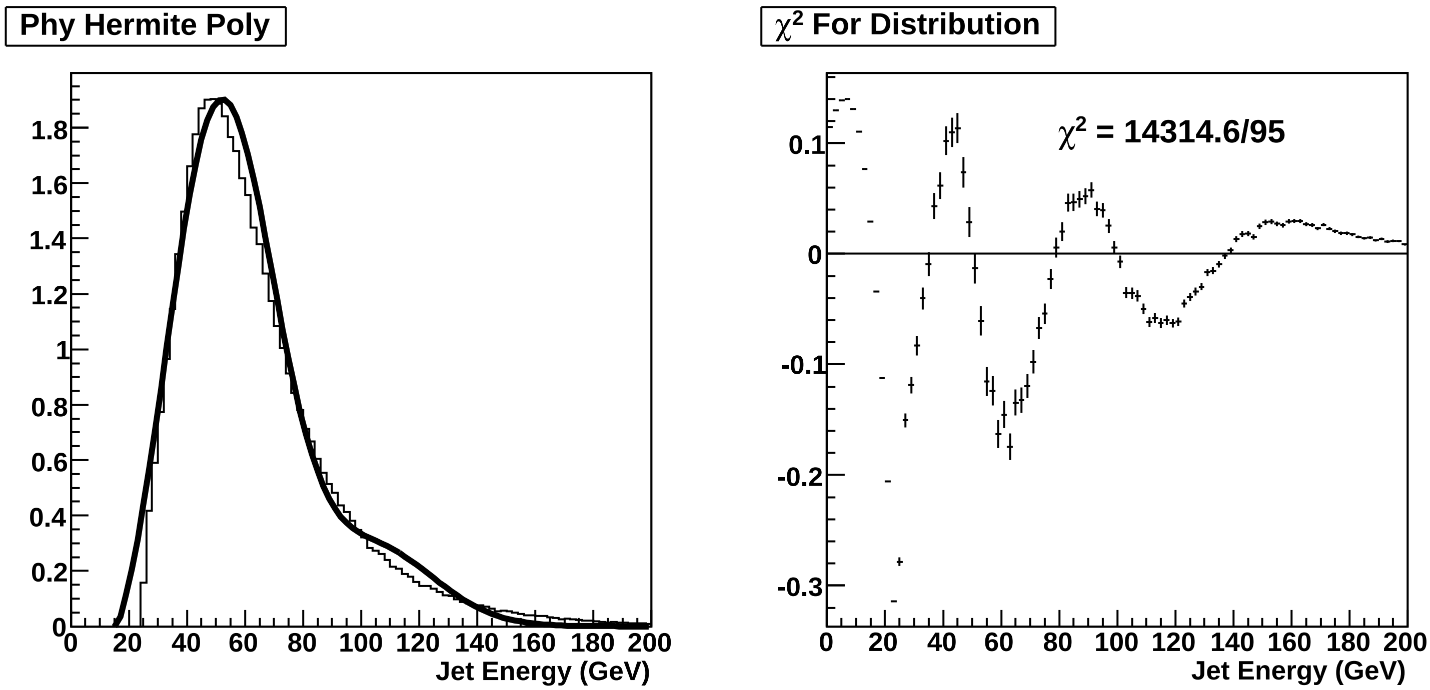

Figure 10: Calculating Probability using Moments and Phys Hermite Poly. Order 5

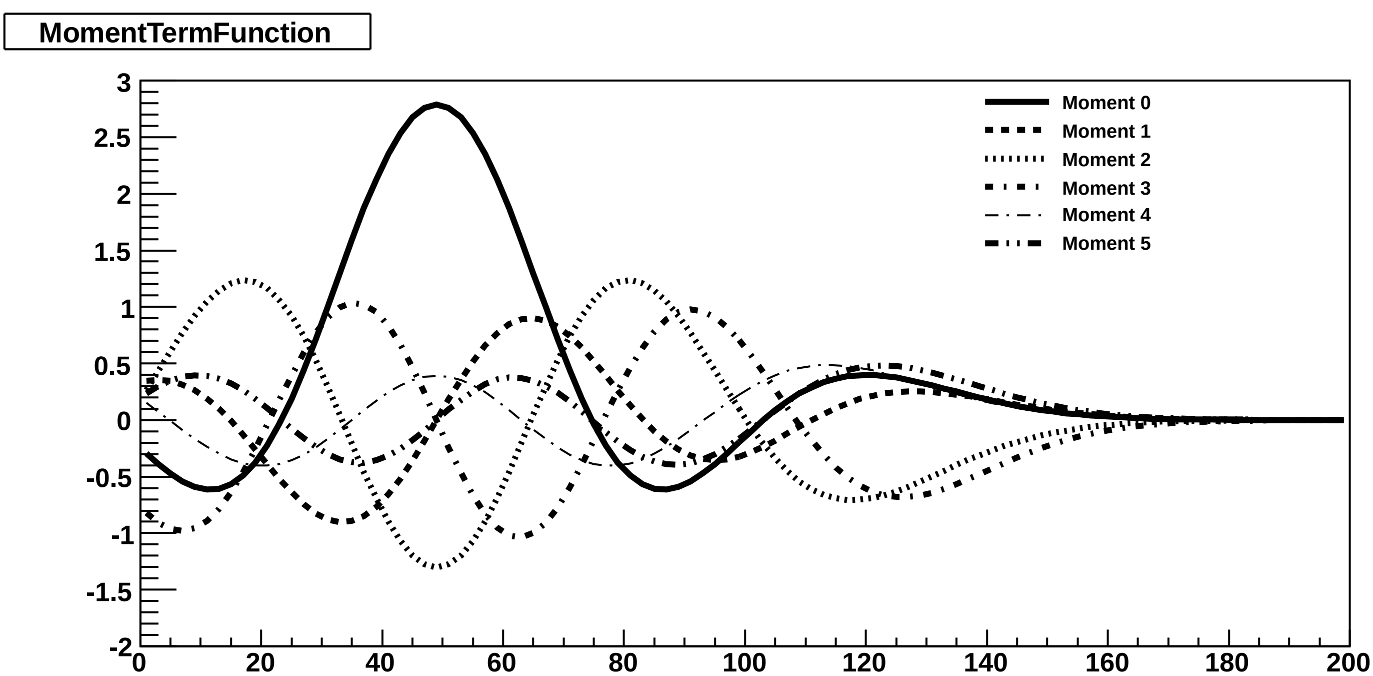

Figure 11: Calculating Probability using Moments and Phys Hermite Poly. Order 5

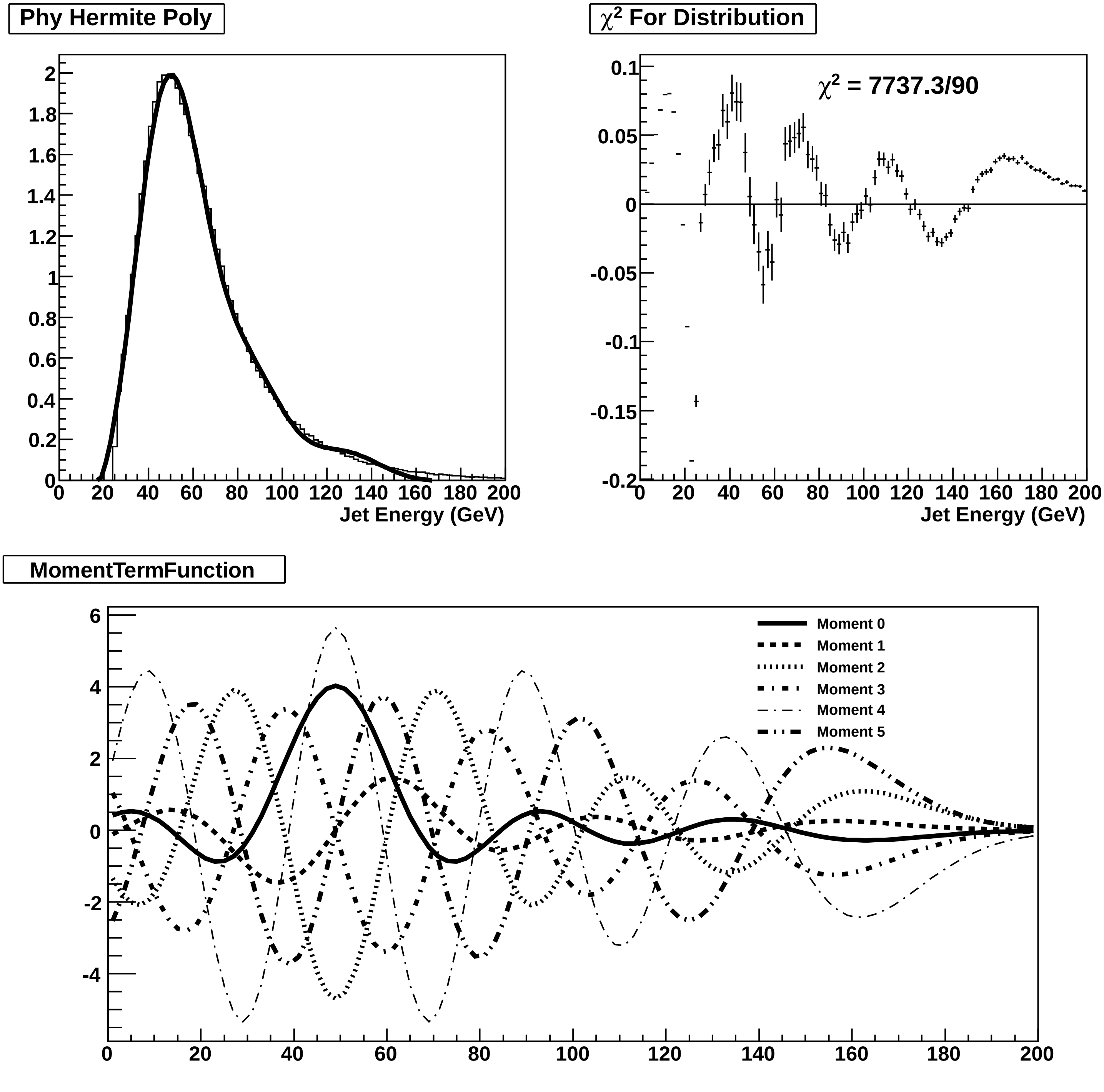

Figure 13: Calculating Probability using Moments and Phys Hermite Poly. Order 10

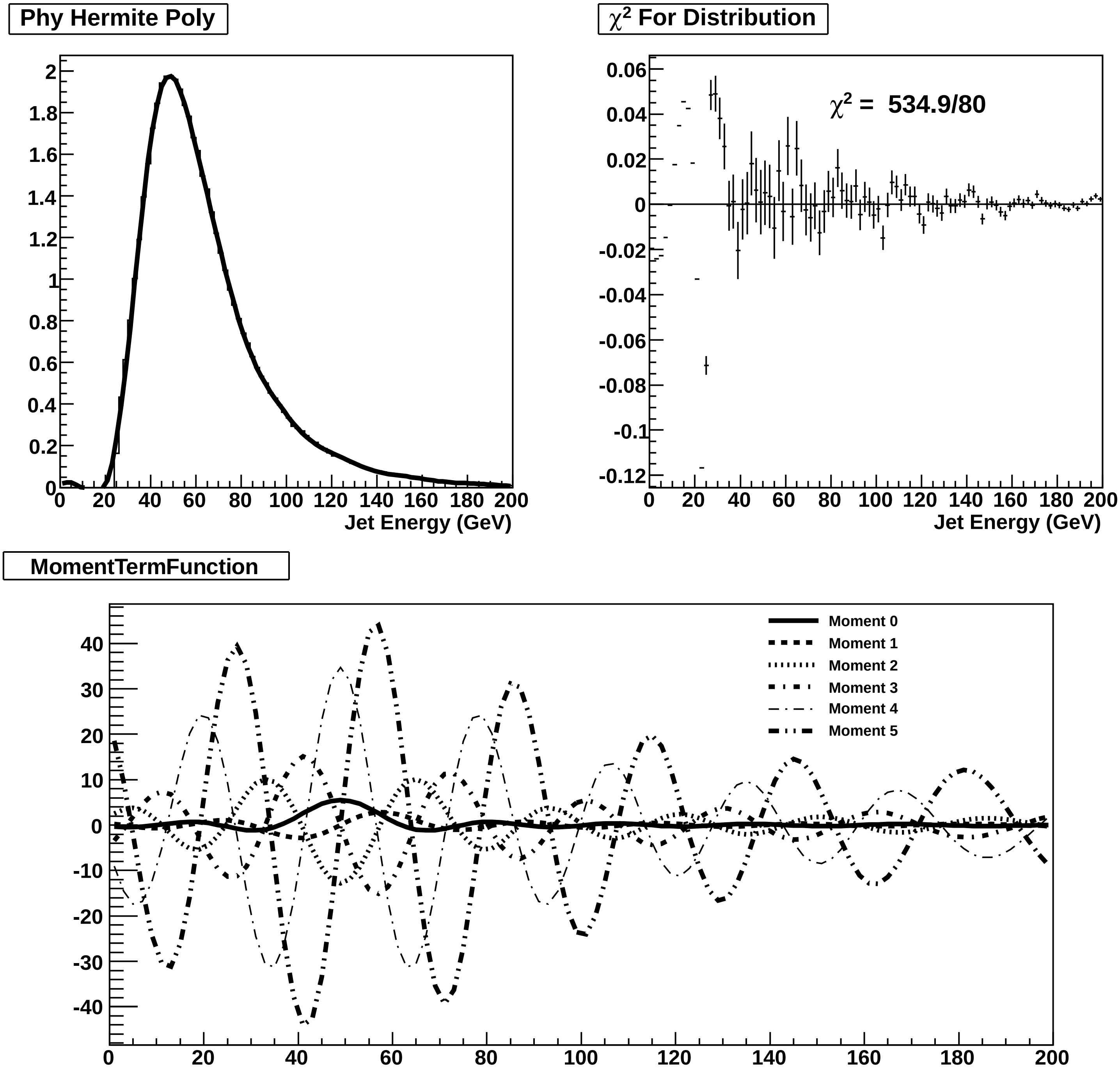

Figure 15: Calculating Probability using Moments and Phys Hermite Poly. Order 20

5 Incorporating Uncertainties Into Fits

The errors on the functions were calculated by assessing the errors on the moments. The errors are very close to the probability distribution and become more accurate as the order increases.

Figures of lower order show the error practically lying on top of the probability function. Some terminology used in these plots is Prob + and Prob -. Prob + refers to the probability function plus the error and Prob- refers to the probability function minus the error.

When the Order became too high, the function went beyond a critical threshold. The function no longer converged and the uncertainties are very chaotic (unfortunately these figures were not provided). At Orders between 20 and 30 the function fit got progressively worse.

$$\delta^2 f = \sum_i \sum_j \left(\frac{\delta f}{\delta \mu_i}\right) \left(\frac{\delta f}{\delta \mu_j}\right) cov(\mu_i, \mu_j) \tag{28}$$

$$\delta^2 f = \left(\frac{\delta f}{\delta \mu_0}\right)^2 \delta^2 \mu_0 + \left(\frac{\delta f}{\delta \mu_1}\right)^2 \delta^2 \mu_1 + 2 \left(\frac{\delta f}{\delta \mu_0}\right) \left(\frac{\delta f}{\delta \mu_1}\right) cov(\mu_0, \mu_1) \tag{29}$$

The errors for these distributions were very chaotic until the covariance between the moments was included in the calculation, Eq.(28) and Eq.(29).The plots for moments were provided earlier in this paper as reference. The Moments at Order 10, Fig(7), show that the fluctuations occur to the left of the peak.

Comparing trends between methods shows that the Gauss-Hermite Expansion method is more precise.When comparing different Orders, the distribution converges more accurately at higher orders.

5.1 Uncertainties (Gram-Charlier Type A Expansion)

The calculations of the error takes into account the moments and the correlations between them. The product is a very small error. This following plots are of this analysis using the Gram-Charlier Type A Expansion.

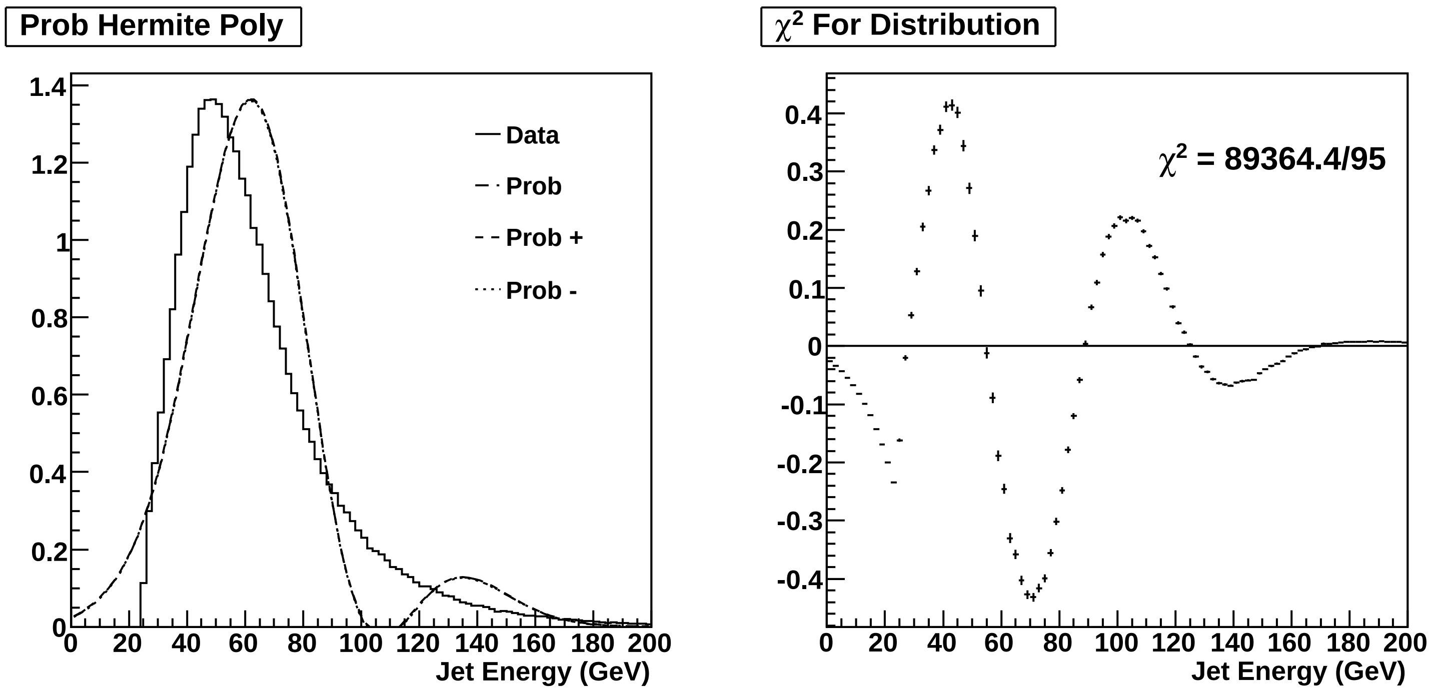

Figure 16: Calculating Error on Prob Fit. Order 5

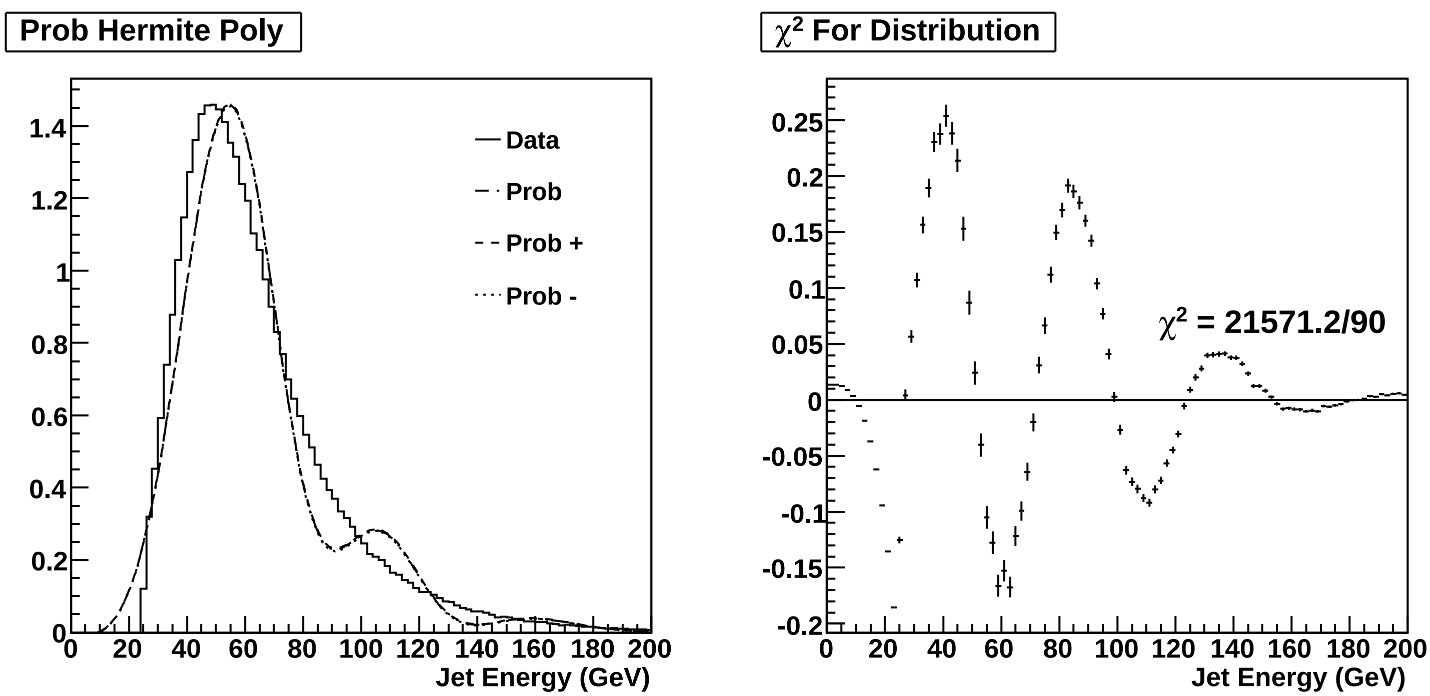

Figure 17: Calculating Error on Prob Fit. Order 10

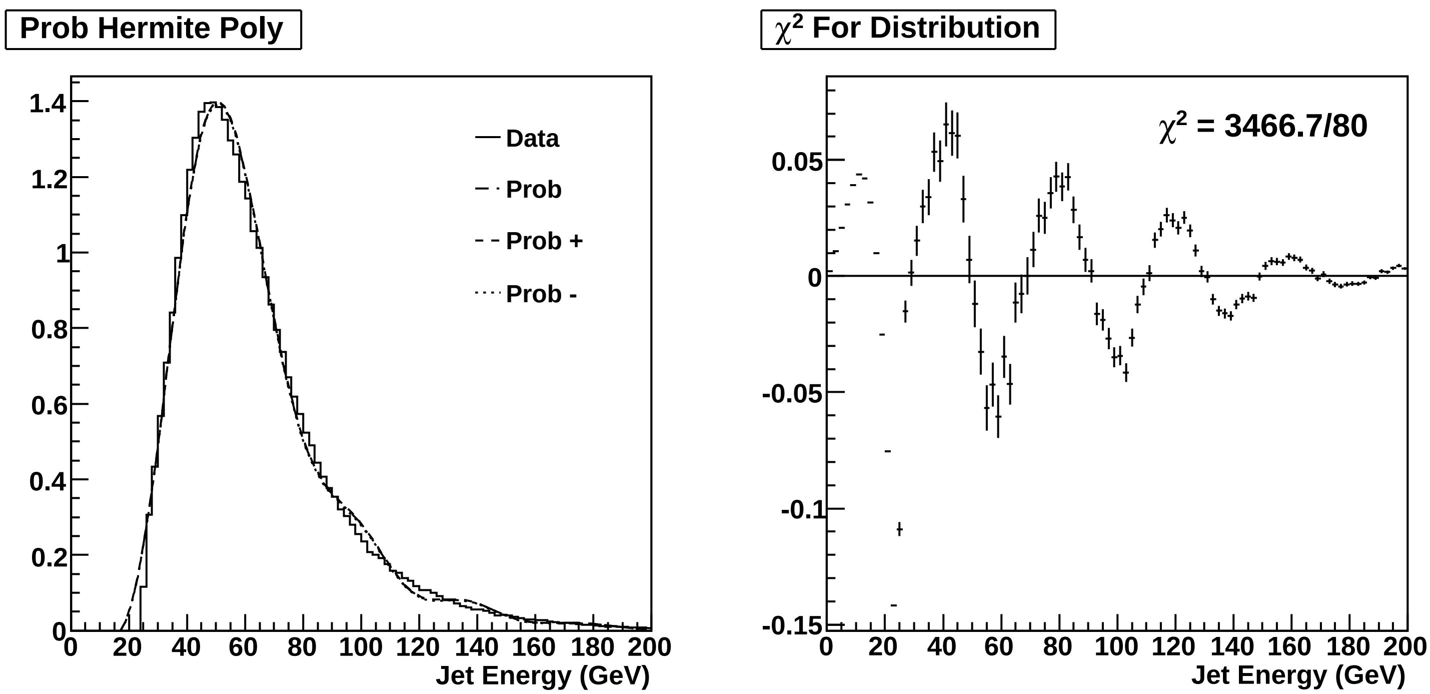

Figure 18: Calculating Error on Prob Fit. Order 20

5.2 Uncertainties (Gauss-Hermite Expansion)

The errors were propagated using the Gauss-Hermite Expansions. In these plots the errors were small and so it seems as though the error function lies ontop of the probability function.

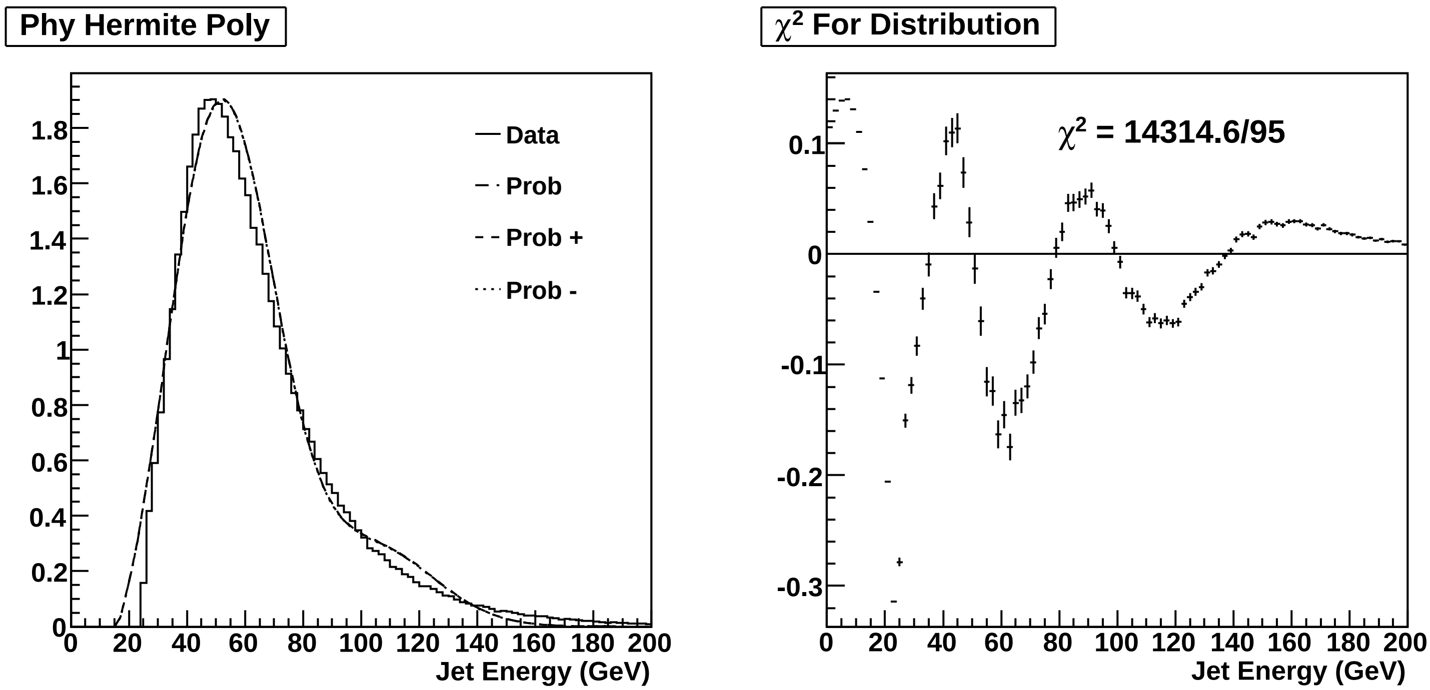

Figure 19: Calculating Error on Phys Fit. Order 5

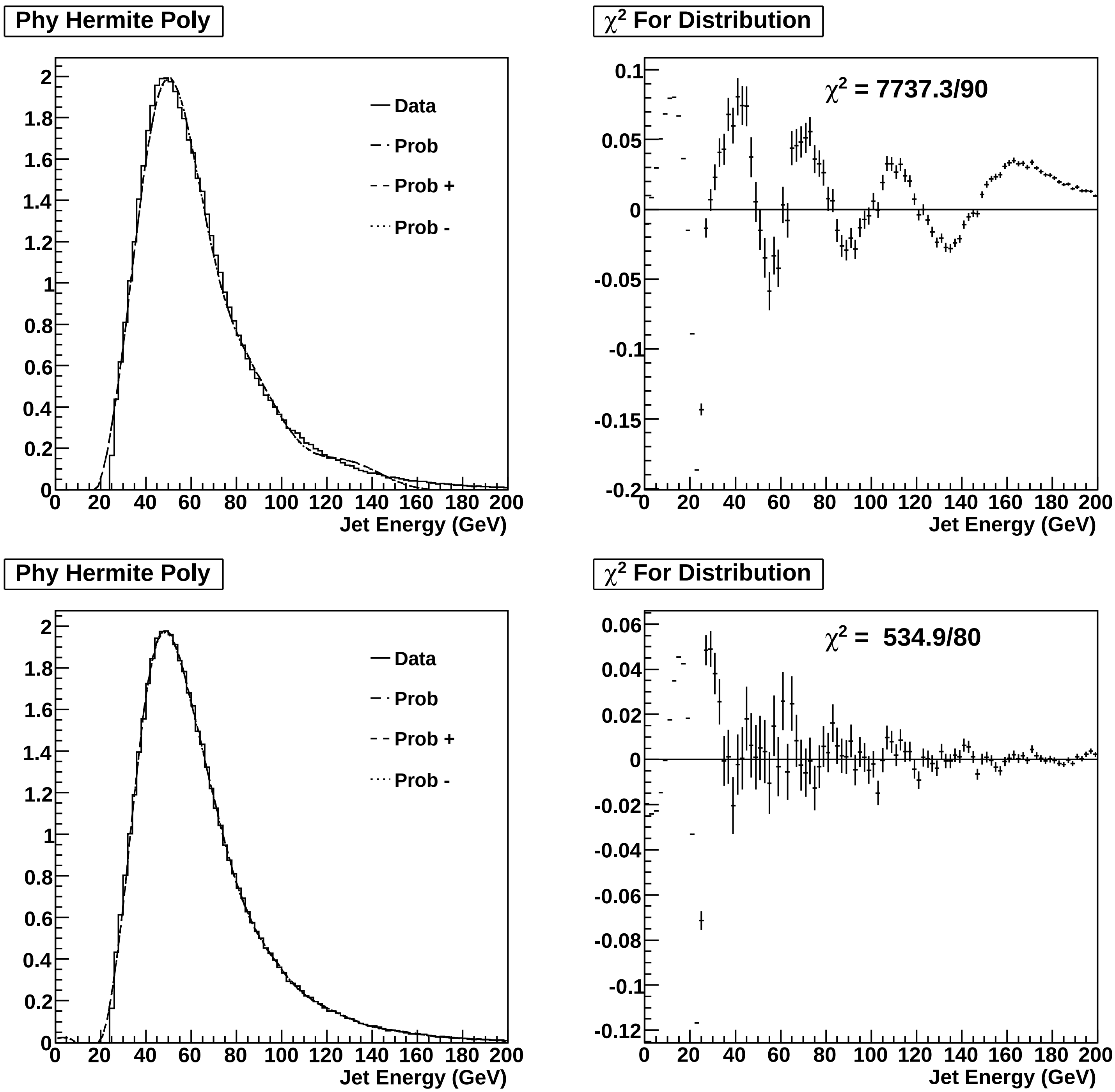

Figure 21: Calculating Error on Phys Fit. Order 20

6 Conclusions

The analysis of the Jetenergy1 distribution was a method of comparing Gram-Charlier Type A and Gauss-Hermite Expansions to find the most accurate fit. It began with formulating specific equations for the Gram-Charlier Type A and the Gauss-Hermite probability functions. This was because both definitions used different Moment terms.

In calculating the moments for the Gauss-Hermite Expansion, μ^z was introduced. The moments were used in calculating the cumulants.

The stages of this analysis continued with searching for the highest order for of the polynomials. The Gauss-Hermite Expansion had polynomials which converged much more quickly than the Gram-Charlier and the Gauss-Hermite Expansion was also able to go to high orders without large uncertainties.

Following this discovery was a calculation of χ^2. The best χ^2 functions were found using the Gauss-Hermite definition. The probability functions for the Gram-Charlier and then the Gauss-Hermite Expansions are not so different, however, the Gauss-Hermite curve fits the distribution much better. The uncertainties were very small for both methods when they were calculated.

The most accurate measurement from this analysis was at Order 20 using the Gauss-Hermite Polynomials. The χ^2 is approximately 6.68 for Order 20 of the Gauss-Hermite definition, the reason it is not closer to 0 is because on the x-axis of the plot the JetEnergy1 distribution began at x=25, while the probability function began at x=0. This led to some discrepancy in the beginning of the plot.

Once these calculations were established on JetEnergy1, Figure(1), a different distribution was attempted. JetEnergy2, Figure(2), had less events but nonetheless it was attempted. It was found to not converge using the Gram-Charlier method but it did converge using the Gauss-Hermite method.

Both JetEnergy1 and JetEnergy2 were used to calculate JetEnergyHT, Figure(3). This was done to see how well the probability function could mimic a distribution when the distribution was not available. This was successful but it could be more precise.

A Definitions

Some terminology specific to this note.

Order : refers to the order of the Hermite Polynomials (will be the same value as the Order for the Moments).
Prob + : The probability distribution plus some small error.
Prob - : The probability distribution minus some small error.

References

[1] S.Blinnikov and R.Moessner, *Expansios for nearly Gaussian Distributions*, Astronomy & Astrophysics Supplement Series (1997), pp 193–205.

[2] M.Welling, *RobustSeries Expansions for Probability Density Estimation*, Dept of Electrical Engineering, California Institute of Technology, pp 136–93.