Likelihood Fitting

CMUPWA-013.v1.0
Gluex-doc-666

Curtis A. Meyer
Carnegie Mellon University

July 15, 2006

Abstract

This note is a short survey of likelihood methods used in partial wave analysis followed by a longer discussion about using the resulting log-likelihoods as goodness-of-fit measures. While there are several approaches that can be used to give a goodness-of-fit measure, most of them suffer in the lack of an absolute calibration on the goodness. Rather, they provide a relative difference, and some guidance on whether the absolute value might be close to the best value.

1 Introduction

This document is an overview of unbinned likelihood fitting. It starts out with a discussion of the standard likelihood method and the generalized likelihood method, then goes on to calculate various relations between them. The next section introduces 1- and 2-D toy Monte Carlos to explore the behavior of the likelihood fits as well as the sensitivities to signals in data sets. Following this, we discuss goodness-of-fit measures.

2 Unbinned Likelihood Fitting

Assume that we have measured N_d events described by $\vec{x}_i^d$ and that we have generated N_r Monte Carlo events thrown uniformly in the appropriate phase space. We describe these events as $\vec{x}_j^m$. Each of the N_r Monte Carlo events is sent through the detector simulation, which yields a sample of N_a accepted Monte Carlo events. We can write an acceptance for each event, η_j, where $\eta_j = 1$ if the event makes it through the Monte Carlo, and is 0 if the event is not accepted. With these events, we can test a physics hypothesis, $I_i(\vec{\alpha}, \vec{x}_i)$, where $\vec{\alpha}$ are parameters that we will optimize. We can define the likelihood as the overall probability of each measured event coming from the parent distribution described by I.

2.1 The Standard Likelihood Function

The standard likelihood function is given as in equation 1 where p is a probability function.

$$L = \Pi_{i=1}^{N_d} p(\vec{x}_i^d) \tag{1}$$

We can build p from our physics hypothesis I by normalizing I over all phase space. We define

$$p(\vec{\alpha}, \vec{x}) = \frac{I(\vec{\alpha}, \vec{x}_i)\eta_i}{\int I(\vec{\alpha}, \tau)\eta(\tau)d\tau} \tag{2}$$

The normalization integral can be computed via Monte Carlo techniques. If we first define the total phase space as

$$\Omega_{tot} = \int d\tau, \tag{3}$$

then the normalization integral can be computed as

$$\int I(\vec{\alpha}, \tau)\eta(\tau)d\tau = \frac{\Omega_{tot}}{N_r} \sum_{j}^{N_r} I(\vec{\alpha}, \vec{x}_j^r)\eta_j \,. \tag{4}$$

Using the fact that η is 1 for accepted events, and zero otherwise, we can rewrite equation 4 as a sum over the accepted Monte Carlo as in equation 5.

$$\int I(\vec{\alpha}, \tau)\eta(\tau)d\tau = \frac{\Omega_{tot}}{N_r} \sum_{j}^{N_a} I(\vec{\alpha}, \vec{x}_j^a) \,. \tag{5}$$

If we now assume that the acceptance of each data event is also 1, then we can rewrite the standard likelihood function in equation 1 as:

$$L = \Pi_{i=1}^{N_d} \frac{I(\vec{\alpha}, \vec{x}_i)}{\frac{\Omega_{tot}}{N_r} \sum_j^{N_a} I(\vec{\alpha}, \vec{x}_j^a)} \tag{6}$$

If we now take the negative of the natural logarithm of equation 6, we get the expression:

$$-\ln(L) \quad = \quad N_d \ln\left[\frac{\Omega_{tot}}{N_r}\sum_j^{N_a} I(\vec{\alpha},\vec{x}^a_j)\right] - \sum_i^{N_d} \ln\left[I(\vec{\alpha},\vec{x}_i)\right] \tag{7}$$

The maximum likelihood procedure now involves minimizing equation 7 with respect to $\vec{\alpha}$. The first part of the sum is known as the normalization integral, and can often be precomputed. The second term is the expensive part of the likelihood fitting procedure.

It should be noted that the standard likelihood function will get the correct shape, but it is not necessarily constrained to get the normalization correct. In particular, if we have chosen things such that the normalization integral is equal to 1, then the negative log likelihood is just given as

$$-\ln(L) \quad = \quad -\sum_i^{N_d} \ln\left[I(\vec{\alpha},\vec{x}_i)\right] \tag{8}$$

When we sum over the accepted Monte Carlo, we will get 1, rather than the expected number of data events, N_d. This means that we need to scale the individual weights by a factor of N_d divided by the normalization integral in equation 7.

2.2 The Generalized Likelihood Function

In the case where we have measured N_d events and can write an expected number of events $< n >$ as a function of the fit parameters, $\vec{\alpha}$, then we can write the likelihood from equation 1 as in equation 9.

$$L \quad = \quad \left(\frac{< n >^{N_d}}{N_d!} e^{-<n>}\right) \cdot \Pi_i^{N_d} \frac{I(\vec{\alpha},\vec{x}^d_i)}{\int I(\vec{\alpha},\tau)\eta(\tau)d\tau} \,. \tag{9}$$

This is known as the generalized likelihood function and describes the joint probability of observing N_d events when $< n >$ are expected and that the observed events are described by the distribution function, $I(\vec{\alpha},\vec{x})$. The integral in the denominator is taken over all phase space and is used to normalize the intensity I. It is convenient to write this as:

$$\int I(\vec{\alpha},\tau)\eta(\tau)d\tau \quad = \quad \frac{\Omega_{tot}}{N_r} \cdot \sum_j^{N_r} I(\vec{\alpha},\vec{x}^r_j)\eta_j \tag{10}$$

where we expect that the integral should be equal to the expected number of events, $<n>$. The sum in equation 10 can be changed to one over the accepted events to give that:

$$<n> \;=\; \frac{\Omega_{tot}}{N_r}\sum_j^{N_a} I(\vec{x}_j^a) \tag{11}$$

which can be used to rewrite equation 9 as equation 12.

$$L \;=\; \frac{1}{N_d!}e^{-<n>}\Pi_i^{N_d} I(\vec{\alpha},\vec{x}_i^d) \tag{12}$$

Taking the natural logarithm of both sides, and negating, we arrive at the following expression for the negative log-likelihood.

$$-\ln(L) \;=\; \ln(N_d!) + <n> - \sum_i^{N_d}\ln\left[I(\vec{\alpha},\vec{x}_i^d)\right] \tag{13}$$

Using Stirling's approximation for $n!$,

$$n! \;\approx\; n^n e^{-n}\sqrt{2\pi n},$$

it is possible to approximate the natural log of a factorial as

$$\ln(n!) \;\approx\; n\left[\ln(n) - 1\right] + \frac{1}{2}\ln(2\pi n)$$

Replacing $<n>$ with the integral in equation 10 and using Stirling's approximation for the natural log of a factorial, we find that:

$$\begin{aligned} -\ln(L) \;=&\; N_d\left[\ln(N_d) - 1\right] + \frac{1}{2}\ln(2\pi N_d) \\ +&\; \frac{\Omega_{tot}}{N_r}\sum_j^{N_a} I(\vec{\alpha},\vec{x}_j^a) - \sum_i^{N_d}\ln\left[I(\vec{\alpha},\vec{x}_i^d)\right] \end{aligned} \tag{14}$$

The first line just contains numeric constants that only depend on the number of data events, N_d. The first term in the second line is the normalization integral, while the last term is the computationally expensive piece. At convergence, the normalization integral piece is driven to be equal to N_d, which means that the final value of the likelihood is given as:

$$-\ln(L)_{fit} \;=\; N_d\ln(N_d) + \frac{1}{2}\ln\left(2\pi N_d\right) - \sum_{i=1}^{N_d}\ln\left[I(\vec{\alpha},\vec{x}_i^d)\right] .$$

Unlike the standard likelihood function, then generalized likelihood function will guarantee that the Monte Carlo will normalize to the number of data events. Finally, if we were to compute both the standard and general likelihood functions for a data set, we would find that

$$-\ln(L_{standard}) \;=\; -\ln(L_{general}) - \frac{1}{2}\ln\left(2\pi N_d\right) - \left(N_d - 1\right). \qquad (15)$$

2.3 The Phase Space Factor

The phase space factor, Ω_{tot}, that is used in equation 3 is obtained by integrating the Lorentz invariant phase space, $dLips$, over all space.

$$\Omega_{tot} \;=\; \int dLips$$

In the case of θ and ϕ in spherical coordinates, $\Omega_{tot} = 4\pi$. In the case of two-body phase space, the integral can be done, and yields that

$$\Omega_{tot} \;=\; \frac{1}{4\pi}\frac{p}{\sqrt{s}}.$$

For higher dimension space, the integral may need to b done numerically, but in the end, it is just a number.

In a situation where we have data binned in $\sqrt{s}$, it is likely that each bin will have its own Ω_{tot}. In addition, the number of data events, N_d, and the intensity, I are also likely to be functions of the bin.

In the end, while we are interested in fitting the data in each bin, we are more likely to be interested in how more global quantities vary as we go from bin to bin. The most obvious global choice would be to have the fit produce a physical cross section, rather than the number of observed events. This can be accomplished by weighting the normalization integral.

2.4 Corrections to The Log-Likelihood

Let us consider a situation where we use the generalized likelihood method, but have modified the normalization to produce a cross sections rather than the number of counts in the bin. This will cause equation 10 to be modified to yield the normalization condition in equation 16.

$$< n > \;=\; \frac{F_{tot}}{N_r}\sum_{j}^{N_a} I(\vec{x}_j^a). \qquad (16)$$

If we now introduce a parameter, $A_{scale} = F_{tot}/\Omega_{tot}$, then we can define a scaled amplitude $J(\vec{\alpha}, \vec{x})$ such that

$$J(\vec{\alpha}, \vec{x}) = A_{scale} \cdot I(\vec{\alpha}, \vec{x}) .$$

This would yield the normalization condition

$$< n > = \frac{\Omega_{tot}}{N_r} \sum_j^{N_a} J(\vec{x}_j^a) . \tag{17}$$

This transformation will result in equation 13 being rewritten as follows.

$$-\ln(L) = \ln(N_d!) + < n > - \sum_i^{N_d} \ln \left[A_{scale} \cdot J(\vec{\alpha}, \vec{x}_i^d) \right] \tag{18}$$

Which allows us to factor out A_{scale} and obtain:

$$-\ln(L) = \ln(N_d!) + < n > - \sum_i^{N_d} \ln \left[J(\vec{\alpha}, \vec{x}_i^d) \right] - N_d \ln \left(A_{scale} \right) . \tag{19}$$

This will modify equation 14 to yield:

$$\begin{aligned} -\ln(L) &= N_d \left[\ln(N_d) - 1 \right] + \frac{1}{2} \ln(2\pi N_d) - N_d \ln \left(A_{scale} \right) \\ &+ \frac{N_a}{N_r} \frac{\Omega_{tot}}{N_a} \sum_j^{N_a} J(\vec{\alpha}, \vec{x}_j^a) - \sum_i^{N_d} \ln \left[J(\vec{\alpha}, \vec{x}_i^d) \right] . \end{aligned} \tag{20}$$

We can summarize this in the following. If we have a likelihood, L_0 where the normalization is unscaled, ($A_{scale} = 1.0$) and a second likelihood, L_A where the scale factor is not equal to 1, then we can relate these two using the following:

$$\ln(L_0) = \ln(L_A) - N_D \ln(A_{scale}) ; \tag{21}$$

there is just a constant offset between the two methods.

3 The Toy Monte Carlos

3.1 The 1-D Monte Carlo

In order to understand how likelihood fitting works, we have built a simple 1-dimensional toy Monte Carlo and likelihood fitter. A weight function based on the first five Legendre Polynomials is used to both generate a data set and to fit the data set. As a reference, the first five Legendre Polynomials are listed below. Note that these are not normalized to 1, but rather to $2/(2n+1)$, where n is the order of the polynomial.

$$\begin{aligned} L_0(x) &= 1.0 \\ L_1(x) &= x \\ L_2(x) &= \frac{1}{2}\left(3x^2 - 1\right) \\ L_3(x) &= \frac{1}{2}\left(5x^3 - 3x\right) \\ L_4(x) &= \frac{1}{8}\left(35x^4 - 30x^2 + 3\right) \end{aligned}$$

We define an intensity function as in equation 22, where the elements of the $\vec{\alpha}$ are typically of order 1.

$$I(\vec{\alpha}, x) = \left(\sum_{i=0}^{4} \alpha_i L_i(x)\right)^2 \tag{22}$$

In the example test that we performed, we have a weight function in which we have taken the nominal weighting of $\alpha_0 = 1$, $\alpha_1 = 2$ and $\alpha_2 = -1$. Based on this, we have considered several data sets in which the parameter α_3 is set to some number, f. In these studies, f takes on the values 0.00, 0.01, 0.02, 0.05, 0.10, 0.20 and 0.40. In each case, we have generated a data set that consists of 15000 data events which have been thrown according to the weight function (equation 22). In addition, we have produced a common set of 50000 events which have been thrown uniformly in phase space. This latter set serves as the Monte Carlo set. In our notation from earlier, we have that $N_d = 15000$, $N_r = 50000$ and $N_a = 50000$. Finally, we have done, $\Omega_{tot} = 2$.

We have performed likelihood fits to the various data sets with different numbers of Legendre Polynomials (waves) allows. In Table 1 are shown the results for the negative log-likelihood and χ^2 for these fits to the data sets with the given admixture of L_3. The fits were done using the generalized log-likelihood, but converted to a standard log-likelihood using equation 15. We can define the goodness-of-fit by projecting both the data and weighted Monte Carlo into a 100-bin histogram from $x = -1$ to $x = 1$, and computing

N_p	f	$-\ln(L)$	χ^2	N_{df}	Prob
3	0.00	6064.52	111.12	97	0.1549
4	0.00	6060.91	106.79	96	0.2121
5	0.00	6060.75	106.85	95	0.1910
3	0.01	6079.84	109.85	97	0.1757
4	0.01	6074.64	102.79	96	0.2992
5	0.01	6074.47	102.89	95	0.2724
3	0.02	6090.41	111.14	97	0.1010
4	0.02	6082.77	103.79	96	0.2759
5	0.02	6082.60	104.04	95	0.2469
3	0.05	6151.17	122.62	97	0.0405
4	0.05	6134.98	98.58	96	0.4081
5	0.05	6134.66	98.85	95	0.3729
3	0.10	6207.55	154.08	97	0.0002
4	0.10	6171.91	100.31	96	0.3615
5	0.10	6176.29	104.74	95	0.2322
3	0.20	6359.26	238.11	97	0.0000
4	0.20	6252.37	96.00	96	0.4808
5	0.20	6252.19	95.97	95	0.4529
3	0.40	6697.07	530.67	97	0.0000
4	0.40	6346.00	86.37	96	0.7490
5	0.40	6345.19	85.06	95	0.7579

Table 1: Fits to data sets built out of a standard mixture of L_0, L_1 and L_2 with an admixture of L_3 multiplied by the parameter f. The log-likelihood has been converted to a standard likelihood while the χ^2 has been determined by projecting the results onto a 1-dimensional histogram with 100 bins. The number of degrees of freedom are $N_{df} = 100 - N_p$.

a χ^2 in each bin. We report the sum of all these individual χ^2 in the table, which in turn corresponds to 100 minus the number of fit parameters, N_p, degrees of freedom, N_{df}. This then yields a probability that is calculated from both χ^2 and N_{df}.

By looking at the probabilities, it appears that with the $N_d = 15000$, that we are sensitive to values of f between 0.02 and 0.05. This can also be seen in Tables 2, 3, 4, 5 and 6. In these tables, we show what the true values of the parameters, $\vec{\alpha}$, are as well as the values that come out of the fits with

$N_p = 3$, 4 and 5. The values for the 4 and 5 parameter fits in Table 2 for $f = 0$ show the level of uncertainty in parameters that are supposed to be zero.

Name	True Value	3-Fit Value	4-Fit Value	5-Fit Value
α_0	54.41	54.49	54.48	54.48
α_1	108.82	109.30	109.33	109.31
α_2	-54.41	-53.38	-53.37	-53.36
α_3	0.00		2.55	2.52
α_4	0.00			0.59

Table 2: The results of fitting wave set with only L_0, L_1 and L_2 contributions with 3, 4 and 5 free parameters in the fit.

Name	True Value	3-Fit Value	4-Fit Value	5-Fit Value
α_0	54.41	54.44	54.43	54.43
α_1	108.82	108.35	109.36	109.36
α_2	-54.41	-53.45	-53.42	-53.41
α_3	1.09		3.66	3.66
α_4	0.00			0.60

Table 3: The results of fitting wave set with L_0, L_1 and L_2 and a 0.02 contribution from L_3. Results are shown for 3, 4 and 5 free parameters in the fit.

Name	True Value	3-Fit Value	4-Fit Value	5-Fit Value
α_0	54.41	54.50	54.43	54.44
α_1	108.82	108.25	109.33	109.32
α_2	-54.41	-53.51	-53.44	-53.43
α_3	2.72		5.24	5.27
α_4	0.00			0.85

Table 4: The results of fitting wave set with L_0, L_1 and L_2 and a 0.05 contribution from L_3. Results are shown for 3, 4 and 5 free parameters in the fit.

Name	True Value	3-Fit Value	4-Fit Value	5-Fit Value
α_0	54.41	54.64	54.34	54.34
α_1	108.82	108.99	109.22	109.22
α_2	-54.41	-53.67	-53.34	-53.34
α_3	10.88		13.49	13.49
α_4	0.00			0.64

Table 5: The results of fitting wave set with L_0, L_1 and L_2 and a 0.2 contribution from L_3. Results are shown for 3, 4 and 5 free parameters in the fit.

Name	True Value	3-Fit Value	4-Fit Value	5-Fit Value
α_0	54.40	54.72	54.29	54.29
α_1	108.80	108.76	108.55	108.55
α_2	-54.40	-54.04	-53.29	-53.28
α_3	21.76		24.48	24.48
α_4	0.00			1.35

Table 6: The results of fitting wave set with L_0, L_1 and L_2 and a 0.4 contribution from L_3. Results are shown for 3, 4 and 5 free parameters in the fit.

3.2 The 2-D Monte Carlo

We have also generated a 2-dimensional Monte Carlo in the spherical polar angles θ and ϕ. In this case, we write the waves in terms of the standard spherical harmonics, Y_{lm} written as a function of $x = \cos\theta$ and ϕ. In this example, we have used the harmonics up through $l = 2$. Note that $\sin\theta = \sqrt{1-x^2}$.

$$
\begin{aligned}
Y_{00}(x,\phi) &= \sqrt{\frac{1}{4\pi}} \\
Y_{10}(x,\phi) &= \sqrt{\frac{3}{4\pi}}\, x \\
Y_{11}(x,\phi) &= -\sqrt{\frac{3}{8\pi}}\,\sqrt{1-x^2}e^{i\phi} \\
Y_{1-1}(x,\phi) &= \sqrt{\frac{3}{8\pi}}\,\sqrt{1-x^2}e^{-i\phi} \\
Y_{20}(x,\phi) &= \sqrt{\frac{5}{16\pi}}\,\left(3x^2-1\right) \\
Y_{22}(x,\phi) &= \sqrt{\frac{15}{32\pi}}\,\left(1-x^2\right)e^{2i\phi} \\
Y_{21}(x,\phi) &= -\sqrt{\frac{15}{8\pi}}\,\left(x\sqrt{1-x^2}\right)e^{i\phi} \\
Y_{2-1}(x,\phi) &= \sqrt{\frac{15}{8\pi}}\,\left(x\sqrt{1-x^2}\right)e^{-i\phi} \\
Y_{2-2}(x,\phi) &= \sqrt{\frac{15}{32\pi}}\,\left(1-x^2\right)e^{-2i\phi}
\end{aligned}
$$

In order to exercise the likelihood fits, a simple toy Monte Carlo and fitter have been built based on the intensity function given in equation 23, where the elements of the $\vec{\alpha}$ are typically of order 1.

$$
\begin{aligned}
I(\vec{\alpha},x) &= |\alpha_0 Y_{0,0}(x,\phi) & (23)\\
&+ \alpha_1 Y_{1,0}(x,\phi) & (24)\\
&+ \alpha_2 Y_{1,1}(x,\phi) & (25)\\
&+ \alpha_3 Y_{1,-1}(x,\phi) & (26)
\end{aligned}
$$

$$+ \quad \alpha_4 Y_{2,0}(x,\phi) \tag{27}$$
$$+ \quad \alpha_5 Y_{2,2}(x,\phi) \tag{28}$$
$$+ \quad \alpha_6 Y_{2,1}(x,\phi) \tag{29}$$
$$+ \quad \alpha_7 Y_{2,-1}(x,\phi) \tag{30}$$
$$+ \quad \alpha_8 Y_{2,-2}(x,\phi)|^2 \tag{31}$$

In the example test that we performed, we have a nominal weight function in which we have taken the nominal weighting as in Table 7. In these studies, f takes on the values 0.00, 0.10, 0.20 and 0.40. For each value of

α_0	1.00
α_1	2.00
α_2	0.00
α_3	-2.00
α_4	-2.00
α_5	0.00
α_6	0.00
α_7	0.00
α_8	f

Table 7: The nominal weighting for the 2-D toy Monte Carlo. The value of f is varied from 0.0 to 0.8.

f, we have generated a data set consisting of 15000 events which have been thrown according to the weight function (equation 23). In addition, we have produced a sample of 50000 events which have been thrown uniformly in phase space. This latter set serves as the Monte Carlo set. This yields that $N_d = 15000$, $N_r = 50000$. With 100% acceptance, $N_a = 50000$ as well. In this Monte Carlo, we have that $\Omega_{tot} = 4\pi$. Assuming that the cross terms in equation 23 do not contribute, then Table 8 gives the fraction of the data set that each value of f represents.

f	0.00	0.10	0.20	0.40	0.80
Events	0	12	49	196	782
Fraction	0.000	0.001	0.004	0.016	0.056

Table 8: The number of events out, and what fraction of the total data set this represents for the values of f used in the 2-D Monte Carlo.

In Tables 9, 10 and 11 are shown results of several fits. Table 9 has $f = 0$ and shows the resulting fit parameters, $\vec{\alpha}$ as well as the negative $\ln(L)$ for fits in which we have 5, 6 and 7 free parameters. In this particular case, the 5-parameter fit should accurately describe the data. It can be seen that the $\ln(L)$ does not significantly improve as additional parameters are allowed, but the fit does allow for a small contribution from the additional waves.

Name	True Value	5-Fit Value	6-Fit Value	7-Fit Value
α_0	90.00	87.35	85.61	86.38
α_1	180.00	177.51	176.72	176.32
α_2				-2.78
α_3	-180.00	-179.14	-180.02	-180.30
α_4	-90.00	-91.00	-90.80	-91.30
α_5	90.00	87.26	88.10	87.55
α_6				
α_7				
α_8			-4.41	-4.20
$-\ln(L)$		33707.54	33704.8	33703.50

Table 9: The results of fitting the data generated with $f = 0$, but fit with an extra parameter, α_8 free.

In Table 10 are shown the results for a $f = 0.10$. In this case, there is also not a significant change in the $\ln(L)$ as additional waves beyond the nominal 5 are allowed. In particular, the actual solution with 6 waves appears no better than that with 5 waves. In this particular case, we do not appear to have any sensitivity the the $f = 0.10$ signal. It is however interesting that the difference between the α_8 contribution to in this case, and that shown if Table 9 is about the right size. There is apparently some underlying fluctuation in the data that is described by the negative value in the $f = 0$ fit.

In Table 11 are shown the results for the $f = 0.20$ case. Here, we start to see some improvement in the $\ln(L)$ as the sixth wave is added into the fit. As a check to understand if the Monte Carlo data set is large enough, we doubled N_r to 100000 events. These results are shown in the lower half of the table. In particular, the likelihood values have changed, the the contributions of the various waves are independent of the number of Monte Carlo events.

Name	True Value	5-Fit Value	6-Fit Value	7-Fit Value
α_0	90.00	84.22	85.76	86.80
α_1	180.00	176.57	177.26	176.64
α_2				-4.00
α_3	-180.00	-180.17	-179.35	-180.81
α_4	-90.00	-90.26	-90.45	-91.12
α_5	90.00	89.06	88.29	87.53
α_8	9.00		4.03	4.22
$-\ln(L)$		33861.67	33859.33	33856.62

Table 10: The results of fitting wave set with Y_{00}, Y_{1m} and Y_{2m} and a 0.10 contribution fromL_3. Results are shown for 5, 6 and 7 free parameters in the fit.

In Table 12 are show the $\ln(L)$ values for the six wave fit with 50000, 100000 and 200000 Monte Carlo events. While the exact value of the likelihood does move around, the fit parameters are stable against this change. This leads us to believe that in the case of 15000 data events, a Monte Carlo sample that is about three times larger is sufficient.

Name	True Value	5-Fit Value	6-Fit Value	7-Fit Value
α_0	90.00	80.44	84.95	85.82
α_1	180.00	175.21	177.14	176.58
α_2				-3.43
α_3	-180.00	-181.39	-179.94	-179.38
α_4	-90.00	-90.15	-90.67	-91.22
α_5	90.00	90.75	88.68	88.05
α_8	18.00		12.15	12.26
$-\ln(L)$		33938.87	33917.43	33915.41
α_0	90.00	81.03	85.54	86.12
α_1	180.00	175.16	177.06	176.68
α_2				-2.32
α_3	-180.00	-181.45	-179.06	-179.35
α_4	-90.00	-90.04	-90.56	-90.93
α_5	90.00	91.39	89.07	88.64
α_8	18.00		12.22	12.32
$-\ln(L)$		33902.19	33880.37	33878.45

Table 11: The results of fitting wave set with Y_{00}, Y_{1m} and Y_{2m} and a $f = 0.20$ contribution from the α_8 wave. Results are shown for 5, 6 and 7 free parameters in the fit. The upper part of the table has 50000 Monte Carlo events, while the lower part of the table has 100000 events. The fit results are not sensitive to the size of the Monte Carlo sample.

Monte Carlo	50000	100000	200000
$-\ln(L)$ (Standard)	33917.43	33880.37	33936.91
$-\ln(L)$ (Generalized)			33936.90

Table 12: The $\ln(L)$ of a good fit as the number of Monte Carlo events is varied. All three of these fits yield the same fit parameters.

4 Goodness of Fit Measures

One of the issues that makes unbinned likelihood fits difficult to use is the fact that the log-likelihood itself is not a goodness-of-fit estimator. It can be shown that under some fairly broad conditions, twice the change in the log-likelihood is equal to the change is χ^2 of a fit. However, in order to be able to use this, the same data set has to be used to compute all log-likelihoods.In a situation where one would like to compare several different data sets, comparisons between the likelihoods of fits to the various data sets cannot be directly made. Both of these trends can be directly seen in the data presented in Table 1.

4.1 Computing a χ^2 per degree of freedom

When examining 1-dimensional models, it is usually possible to define a χ^2 for a fit. This is accomplished by binning the data in the single variable and then comparing the data and fit result on a bin-by-bin basis. We can define the the total χ^2 as in equation 32.

$$\chi^2 = \sum_{i=1}^{bins} \frac{(d_i - f_i)^2}{\sigma_{d_i}^2 + \sigma_{f_i}^2} \tag{32}$$

In the above formula, d_i are the number of data events in bin i, f_i are the number of fit events in the bin i, σ_{d_i} is the error in the number of data events and σ_{f_i} is the error in the number of fit events. The number of degrees of freedom, N_{df}, is the number of bins for which there is non-zero acceptance minus the number of fit parameters, α. The χ^2 and *ndf* can be used to compute a probability that the data came from the parent distribution described by the fit. Figure 1 shows the mapping between χ^2, N_{df} and probability.

One issue with the formula 32 is the value of the σ's used in the sum. For sufficiently large numbers of data events, d_i, we can use the Gaussian approximation that $\sigma_d = \sqrt{d}$. Similarly, the error on the fit potentially has two components. The first is a statistical term that comes from the size of the Monte Carlo sample. The second is an error in the fit parameters themselves. In fact, these are correlated, and in computing the χ^2, it usually makes sense to use only one or the other when determining the χ^2 of a fit. In this note, we will only consider the statistical error in the Monte Carlo. Under this assumption we can compute the χ^2 as follows. If there are r_i Monte Carlo

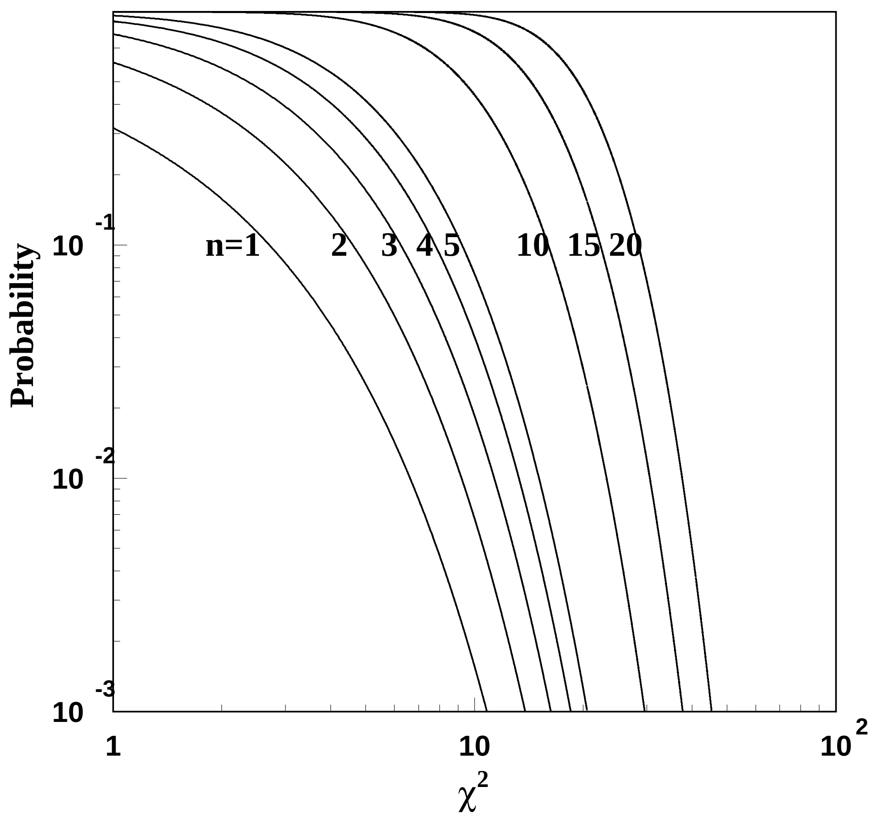

Figure 1: Probability versus χ^2 for different numbers of degrees of freedom, N_{df}.

events in bin i, then the error on f_i due to statistics is:

$$\begin{aligned}\sigma_{f_i} &= \sqrt{r_i}\frac{f_i}{r_i} \\ &= \sqrt{f_i}\sqrt{\frac{f_i}{r_i}}.\end{aligned}$$

Unfortunately, if either d_i of f_i are small, then the above approximations underestimate the errors associated with the small number of events. In order to handle this, one should use Poisson statistics, which then allows us to define an upper and lower error limit, each of which exclude 15.9% of the data. This correcponds to what a 1σ limit with Gaussian errors would do. We refer to these to limits as a_l and a_h, so if there are n counts, then the range between $n - a_l$ to $n + a_h$ has a 62% chance of containing the number of events in the parent sample. This derivation was carried out in reference [1] and the following formula have been taken from there. They can be used to compute the values of a_l and a_h given in Table 13.

$$\begin{aligned} 0.841 &= e^{-a_l} \textstyle\sum_{i=0}^{n-1} \left[\frac{(a_l)^{(i)}}{i!}\right] &= \frac{\Gamma(a_l, n)}{\Gamma(n)} \\ 0.159 &= e^{-a_h} \textstyle\sum_{i=0}^{n} \left[\frac{(a_h)^{(i)}}{i!}\right] &= \frac{\Gamma(a_h, n+1)}{\Gamma(n+1)} \end{aligned}$$

The quantity n is the number of counts in the bin, while $\Gamma(n)$ is the normal gamma function, (just $n!$ for integer n). The function $\Gamma(a, n)$ is the *generalized, regularized incomplete gamma function* [**?**]. These can be inverted, and the following Mathematica functions will evaluate the upper and lower limits.

$$\begin{aligned} a_l &= \text{InverseGammaRegularized}(n, 0.841) \\ a_h &= \text{InverseGammaRegularized}(n+1, 0.159) \end{aligned}$$

In Table 13 are given the Gaussian errors as well as the lower and upper errors evaluated from above. In addition, the last column gives a reasonably good approximation to the upper limit for small numbers. Unfortunately, when we try to use these errors in computing χ^2, we find that these are an overestimation of the errors. In fact, the best approach appears to be to use $\sqrt{k}$ as the error on k events, and in the case where $k = 0$, we simply set the error in k to be 1 rather than the nominal 0. This gives much better values for the χ^2, but slightly larger than they should be. However, they appear much closer than the case where the errors on small numbers are used.

4.2 Likelihood Difference Methods

In Table 1 are presented various fits to seven different data sets, each with the same number of events. While its is possible to compare different fits to the same data sets, (denoted by f value), comparisons of fits to different data sets are not directly possible. The nominal good fits, ($N_p = 4$), to each of the data sets vary from 6061 up to 6346. All of which are equivalently

Count k	Normal Error $\sqrt{k}$	Lower Error $k - a_l$	Upper Error $a_h - k$	Approx Upper $1 + \sqrt{k + 0.75}$
0	0	0	1.84	1.87
1	1	0.83	2.30	2.32
2	1.41	1.29	2.64	2.66
3	1.73	1.63	2.92	2.94
4	2.00	1.91	3.16	3.17
5	2.24	2.16	3.37	3.40
6	2.45	2.38	3.58	3.60
7	2.65	2.58	3.77	3.79
8	2.83	2.77	3.94	3.96
9	3.00	2.94	4.11	4.12
10	3.16	3.11	4.26	4.28
11	3.32	3.26	4.41	4.43
12	3.46	3.41	4.55	4.57
13	3.60	3.56	4.69	4.71
14	3.74	3.69	4.82	4.84
15	3.87	3.83	4.95	4.97
16	4.00	3.95	5.07	5.09
17	4.12	4.08	5.19	5.21
18	4.24	4.20	5.31	5.33
19	4.36	4.31	5.42	5.44
20	4.47	4.47	5.54	5.56
50	7.07	7.05	8.11	8.12
100	10.00	9.97	11.00	11.04
...	...	...	...	...
k	$\sqrt{k}$	$\sqrt{k}$	$\sqrt{k}$	$1 + \sqrt{k + 0.75}$

Table 13: Upper and lower limit errors for small numbers of counts.

good fits. One cannot simply us a log-likelihood value as a goodness-of-fit measure.

Recently there has a been a proposed method to turn the $\ln(L)$ into a goodness-of-fit measure [3]. In the remainder of this section, we examine the feasibility of using methods described in reference [3] to turn a log-likelihood into a goodness of fit measure. This method will be briefly described here.

The basic idea is to first note that ratios of likelihoods, or differences in log-likelihoods can be used as improvement measures for fits to the same data set. This leads to the idea of defining a likelihood measure based solely on the data set being used, and then compare the results of a fit to that of the data likelihood.

The data likelihood is a standard likelihood function, L_d, that is built out of a probability function that can be computed using only the data. In the case where the data are described by a set of observables, $\vec{x}_i$ for N_d events, the data probability function is defined as

$$P^{data}(\vec{x}) \quad = \quad \frac{1}{N_d}\sum_{i=1}^{N_d}\mathcal{G}(\vec{x}-\vec{x}_i) \tag{33}$$

where $\mathcal{G}(\vec{x}-\vec{x}_i)$ is a kernel function of dimension d which is centered about each measured point. The proposed kernel function is a Gaussian kernel, defined as in equation 34.

$$\mathcal{G}(\vec{x}) \quad = \quad \frac{1}{(\sqrt{2\pi}h)^d\sqrt{E}}\cdot e^{\frac{-H^{\alpha\beta}x^\alpha x^\beta}{2h^2}} \tag{34}$$

where the error matrix $E^{\alpha\beta}$ has determinant E and the Hessian matrix, $H^{\alpha\beta}$, is the inverse of $E^{\alpha\beta}$. The error matrix is given as:

$$E^{\alpha\beta} \quad = \quad < x^\alpha x^\beta > - < x^\alpha >< x^\beta > \tag{35}$$

where the averages are taken over the N_d data events. The remaining parameter, h, is known as a *smoothing parameter*. The reference gives some guidance in choosing this, namely

$$h \quad \approx \quad n^{-1/(d+4)}\,.$$

However, choosing an appropriate value of h appears to be the crux of being able to use this procedure.

Assuming that we have chosen some reasonable value of h, then we can then define a likelihood for the data, L_d which is given in equation 36.

$$L_d \quad = \quad \Pi_{i=1}^{N_d}P^{data}(\vec{x}_i) \tag{36}$$

The negative log-likelihood for the data is then given as in equation 37.

$$-\ln(L_d) \quad = \quad -\sum_{i=1}^{N_d}\ln\left[P^{data}(\vec{x}_i)\right] \tag{37}$$

Such functions can be easily defined for the toy Monte Carlo data sets that were generated in the previous sections.

4.3 Comparison to the 1-D Monte Carlo

In order to attempt to understand what is a reasonable value for h, we start with our 1-D toy Monte Carlo. In this case, $d = 1$, and $N_d = 15000$ which nominally yields a value of $h = 0.15$. In Table 14 are shown the values of the data log-likelihood for ten different values of the smoothing parameter h. Values of h which vary from 0.001 up to 0.20. Unfortunately, the resulting data $\ln(L_d)$ appear to vary widely with the value of h.

h	0.20	0.15	0.10	0.075	0.050
$-\ln(L_D)$ (1D)	7406	7054	6721	6562	6404
h	0.015	0.01	0.005	0.002	0.001
$-\ln(L_D)$ (1D)	6127	6054	5922	5636	5185

Table 14: Evaluation of the data log-likelihood as a function of the value of the smoothing parameter, h, for the 1D toy Monte Carlo.

However, if we plot $\ln(L_d)$ as a function of h as shown in Figure 2, we notice that there is an obvious *knee* in the function. In this particular example, the knee occurs around $h = 0.01$. In Table 15 are shown the results of the fits to the various f data sets as well as $\ln(L_d)$ for for $h = 0.010$ and $h = 0.015$. We note that the values of these two data-log-likelihoods are close to values that are obtained when the fits are good.

The fits given in Table 15 are plotted in Figure 3. The upper plot is χ^2/ndf plotted against the probability of the fit waves describing the data set. As one expects, this is a fairly flat function over most probabilities, with a rapid upward spike in χ^2 for improbable fits. The lower plot is the difference between the log-likelihood of the fit data and the data log-likelihood for $h = 0.01$. This function shows a very similar behavior to the χ^2/ndf plot, which indicates that this difference is a goodness-of-fit measure. In Figure 4 are plotted the two measures against each other. There is a clear correlation between the two, and in principle, both could be used as goodness-of-fit measures. Unfortunately, there is not a linear relation between the two, so it would be difficult to turn the log-likelihood difference into a probability as can be done with a χ^2.

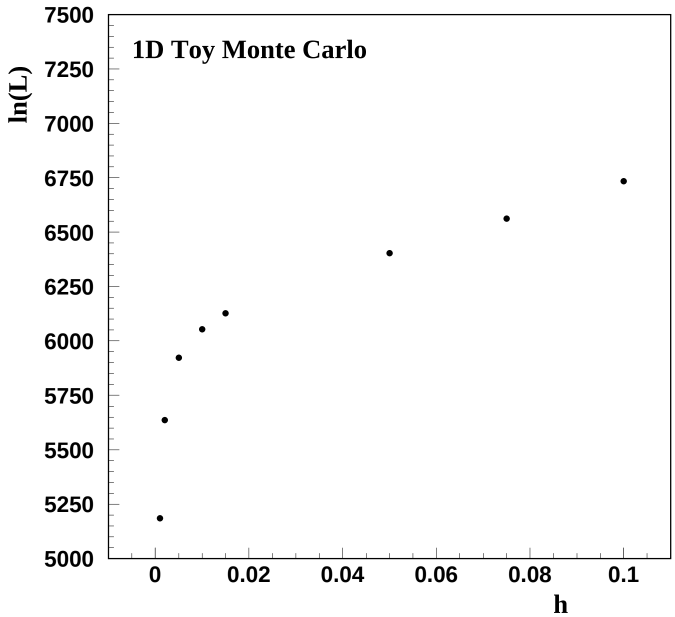

Figure 2: The $\ln(L_d)$ for the 1-D toy Monte Carlo data set as a function of the value of the smoothing parameter, h.

N_p	f	$-\ln(L)$	$-\ln(L_D)$		Prob
			$h = 0.01$	$h = 0.015$	
3	0.00	6064.52	6054.12	6137.26	0.1549
4	0.00	6060.91	6054.13	6127.36	0.2121
5	0.00	6060.75	6054.12	6127.36	0.1910
3	0.01	6079.84	6071.88	6144.72	0.1757
4	0.01	6074.64	6071.88	6144.72	0.2992
5	0.01	6074.47	6071.88	6144.72	0.2724
3	0.02	6090.41	6081.51	6154.14	0.1010
4	0.02	6082.77	6081.51	6154.14	0.2759
5	0.02	6082.60	6081.51	6154.14	0.2469
3	0.05	6151.17	6135.87	6210.47	0.0405
4	0.05	6134.98	6135.87	6210.47	0.4081
5	0.05	6134.66	6135.87	6210.47	0.3729
3	0.10	6207.55	6178.60	6252.16	0.0002
4	0.10	6171.91	6178.60	6252.16	0.3615
5	0.10	6176.29	6178.60	6252.16	0.2322
3	0.20	6359.26	6274.48	6349.56	0.0000
4	0.20	6252.37	6274.48	6349.56	0.4808
5	0.20	6252.19	6274.48	6349.56	0.4529
3	0.40	6697.07	6384.94	6479.03	0.0000
4	0.40	6346.00	6384.94	6479.03	0.7490
5	0.40	6345.19	6384.94	6479.03	0.7579

Table 15: A comparison of the log-likelihoods from the various fits and that computed from the data log-likelihood for two different smoothing parameters for the 1D toy Monte Carlo.

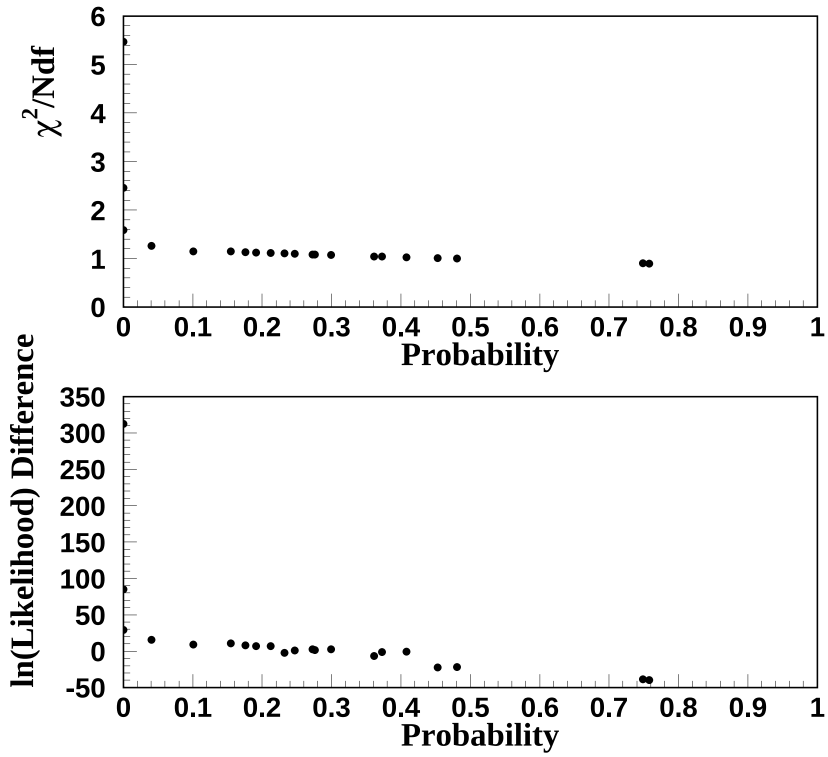

Figure 3: A comparison of the various goodness of fit measures with the computed probability. The upper figure shows the χ^2/ndf versus the probability. The roughly flat line around 1 per degree of freedom corresponds to good fits. As this starts to rise, the probability of the fit goes to zero. The lower plot shows the difference between the fit $\ln(L)$ and the data $\ln(L)$ plotted against the probability. This curve shows a similar shape to the upper plot. It is possible to define a range of likelihood differences that correspond to a good fit in this case.

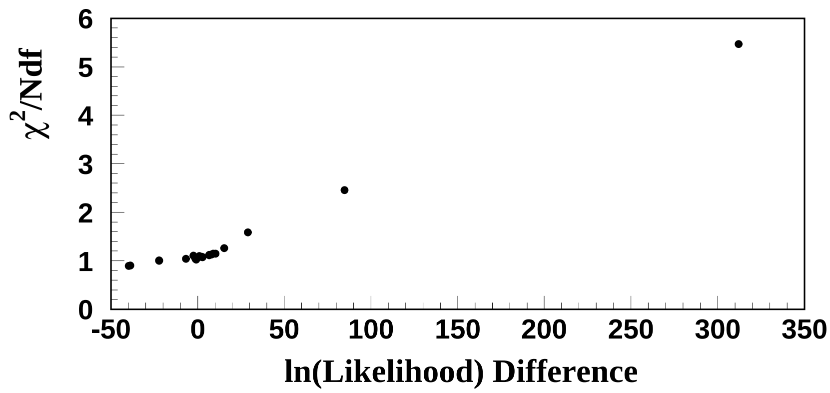

Figure 4: The χ^2/ndf plotted against the likelihood difference. It is possible to define a range of likelihood differences that correspond to a good fit in this case.

4.4 Comparison to the 2-D Monte Carlo

In Table 16 are shown χ^2 as calculated for several different 2-D Monte Carlo sets. The parameter f indicates how much of the Y_{2-1} spherical harmonic is added into the sample. Assuming that there is no interference, then for the various values of f, the following table lists the number of events and the fraction of the total signal. Looking at the results in Table 16, we see that at about 1% of the total signal, we start to have sensitivity to a partial waves contribution. This is comparable to what we would expect from pure statistics on a 15000 event sample. However, it is clear that the changes in the χ^2 for the 2-D $\cos\theta - \phi$ plane do not match well with $2\Delta\ln(L)$. The error are almost certainly not estimated correctly. We also note that in the two cases of 1-D projections, there are also issues with the fit quality not being well measured. This is particularly clear in the case of $\cos\theta$ where none of the fits do a fantastic job in describing the data, while the fits in ϕ all appear to be reasonably good.

Similarly to the 1-D case above, we have examined the data-log-likelihood for several values of the h parameter. These are given in Table 17, and are plotted in Figure 5. As before, there is a clear knee in the data. This time at values of h near about 0.04. These can be compared to the fits to the toy Monte Carlo, as in Table 18. In this particular case, the $N_p = 6$ case correspond to the good fit. Again, there is reasonable agreement between the fit likelihood and the data likelihoods.

N_p	f	$-\ln(L)$	$\cos\theta$			ϕ			$\cos\theta - \phi$		
			χ^2	N_{df}	Prob	χ^2	N_{df}	Prob	χ^2	N_{df}	Prob
5	0.0	33708	114.30	95	0.086	96.58	95	0.436	408.57	395	0.308
6	0.0	33705	112.42	94	0.095	93.75	94	0.488	399.46	394	0.414
7	0.0	33703	111.73	93	0.090	92.14	93	0.506	401.88	393	0.368
5	0.1	33862	116.62	95	0.065	94.25	95	0.503	409.48	395	0.297
6	0.1	33859	115.98	94	0.062	94.35	94	0.471	419.51	394	0.273
7	0.1	33857	114.81	93	0.062	91.46	93	0.523	413.33	393	0.231
5	0.2	33939	126.44	95	0.017	98.30	95	0.388	429.05	395	0.115
6	0.2	33917	115.36	94	0.067	93.83	94	0.485	406.50	394	0.321
7	0.2	33916	114.46	93	0.065	91.20	93	0.533	409.21	393	0.276
5	0.4	34031	150.28	95	0.000	180.74	95	0.000	517.36	395	0.000
6	0.4	33891	106.84	94	0.172	93.50	94	0.495	350.37	394	0.944
7	0.4	33890	105.76	93	0.172	92.89	93	0.484	349.37	393	0.945
5	0.8	34787	403.62	95	0.000	363.31	95	0.000	1231.27	395	0.000
6	0.8	34103	109.37	94	0.133	97.79	94	0.374	396.28	394	0.458
7	0.8	34103	108.64	93	0.128	97.18	93	0.363	396.82	393	0.437

Table 16: Results of fits to the 2-D toy Monte Carlo. The table shows the log-likelihood for each fit, as well as χ^2's calculated for $\cos\theta$, ϕ and the two dimensional plot. Also given are the number of degrees of freedom and the probability for each fit.

h	0.20	0.15	0.10	0.075	0.05
$-\ln(L_D)$ (2D)	37371	36348	35356	34848	34190
h	0.04	0.03	0.020	0.015	0.01
$-\ln(L_D)$ (2D)	33762	33016	31252	29158	24512

Table 17: Evaluation of the data log-likelihood as a function of the value of the smoothing parameter, h, for the 2D toy Monte Carlo.

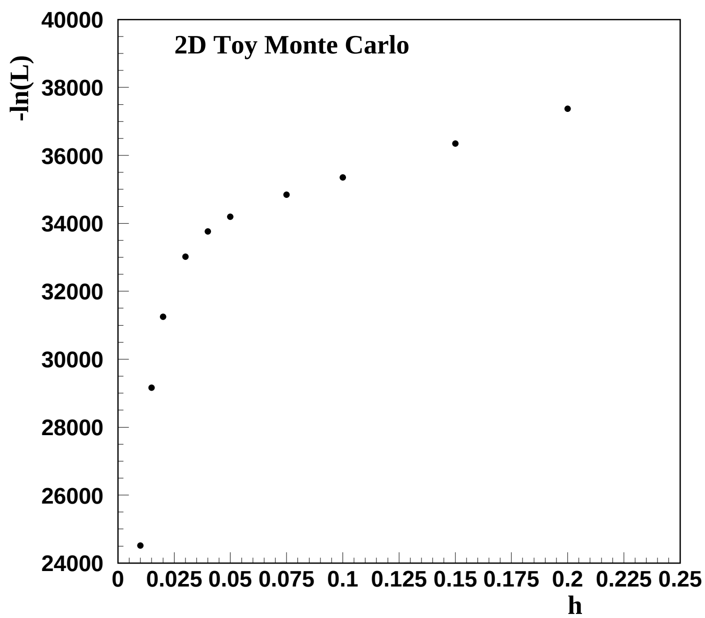

Figure 5: The $\ln(L)$ for the 2-D toy Monte Carlo data sets as a function of the value of the smoothing parameter, h.

N_p	f	$-\ln(L)$		$-\ln(L_D)$	
			$h = 0.05$	$h = 0.04$	$h = 0.03$
5	0.00	33708	34017	33594	32865
6	0.00	33705	34017	33594	32865
7	0.00	33704	34017	33594	32865
5	0.10	33862	34134	33707	32968
6	0.10	33859	34134	33707	32968
7	0.10	33857	34134	33707	32968
5	0.20	33939	34190	33762	33016
6	0.20	33880	34190	33762	33016
7	0.20	33878	34190	33762	33016
5	0.40	34031	34152	33715	32959
6	0.40	33891	34152	33715	32959
7	0.40	33890	34152	33715	32959

Table 18: Comparison of $\ln(L)$ with two goodness of fit measures for the 2D toy Monte Carlo.

4.5 The Mother of All Fits Method

A final approach to turning the log-likelihood into a goodness-of-fit measure is again based on the idea that the likelihood ratio, or the difference in log-likelihoods between two fits to the same data sample corresponds to a change in χ^2 of the fit. Namely that

$$-2\Delta \ln(L) \;=\; \Delta\chi^2 \,.$$

In this particular method, we assume that by giving enough freedom to the fitting function, it will be able to accurately reproduce the data, and yield an *ultimate limit* to the value of $\ln(L)$ for the particular data set. In this way, we can define a *best* log-likelihood, $-\ln(L_{best})$, and then produce the difference:

$$\chi^2_{fit} - \chi^2_{best} \;=\; 2 \cdot [\ln(L_{best}) - \ln(L_{fit})] \,. \tag{38}$$

In order to utilize the measure in equation 38, we need to not that the two χ^2 are computed for different numbers of degrees of freedom, $N_{df}(best)$ and $N_{df}(fit)$, where the value for the best fit is smaller than that for the case of interest by the difference in the number of fit parameters used. In order to utilize this method, we need to also use the data in Figure 1.

5 Summary

We have examined several methods of producing goodness of fits for likelihood fits to data. While it is possible to define many measures that can be used as reaonable goodness-of-fit measures, it is often difficult to turn these into an absolute goodness-of-fit.

References

[1] Victor Regener, *Statistical Significance of Small Samples of Cosmic-Ray Counts* Phys. Rev. **84**, 161, (1951)

[2] Mathematica 5.1, Wolfram Research.

[3] Rajendran Raja, *A General Theory of Goodness of Fit in Likelihood Fits*, physics/0509008 (2005).

[4] S.U. Chung, *Spin Formalisms, Updated Version*, BNLQGS-02-0900, (2004).

[5] A.Anisovich, E.Klempt, A.Sarantsev and U. Thoma, *Partial Wave decomposition of pion and photoproduction amplitudes*, Eur. J. Phys. A**24**, 111, (2005).

[6] A.Anisovich, A.Sarantsev, O. Bartholomy, E. Klempt V.A. Nikonov and U. Thoma, *Photoproduction of baryons decaying into N pi and N eta*, Eur. J. Phys. A**25**, 427, (2005).

[7] A.Anisovich, V.A. Nikonov, A.Sarantsev, E. Klempt and U. Thoma, *Decays of baryon resonances into Lambda K+, Sigma0 K+ and Sigma+ K0* Eur. J. Phys. A**25**, 441, (2005).