CLAS-NOTE 2003-011

A Monte Carlo Study of the Z^+ Peak Significance

Curtis A. Meyer

Carnegie Mellon University

July 23, 2003

Abstract

A Monte Carlo study of the Z^+ peak has been carried out by varying a background distribution according to Poisson statistics and then looking for resulting peaks above background. From this study, I conclude that there is about a 1 in 500,000 chance of the observed peak occurring purely by chance. Under a typical PDG measure, this corresponds to a 4.75σ peak.

A detailed study has been carried out to access the statistical significance of the Z^+ peak observed in the reaction $\gamma d \rightarrow K^+K^-pn$ in CLAS. Monte Carlo methods have been used to determine the probability of a background function statistically fluctuating into a peak. The data used in this study are shown in Figure 1 below. The background shape has been taken to be a triangular function, which is also shown in the figure. The triangle is the intersection of the two lines described as follows:

$$\text{counts} = 113.57 * \text{mass} - 159.73 \qquad (1.40 \leq \text{mass} \leq 1.626) \tag{1}$$

$$\text{counts} = -115.83 * \text{mass} + 213.60 \qquad (1.626 \leq \text{mass} \leq 1.850)\ . \tag{2}$$

In reality, the most important part of the background is that on the rising slope under the Z^+ peak. However, the entire background has been parametrized. The spectrum shown in the figure contains 60 bins of width $10 MeV/c^2$ with the lowest mass edge at $1.405\, GeV/c^2$. The simulated background spectrum used in this analysis has 43 bins, with the center of the lowest mass bin at $m = 1.41\, GeV/c^2$. If we examine the data in Figure 1, we find that the signal is 38.5 counts on top of a background of 45.5

counts in a three bin range centered at $1.542\,MeV/c^2$. The paper reports 43 counts over a background of 54 events in a 3.6 bin wide region centered at the same mass. All analysis in the paper is performed on the three-bin wide sample. The 54 counts in the paper is a good estimate of the background, but the 43 counts in the signal is an overestimate by about $2-3$, which can be seen in the fit that overestimates the peak height by about this.

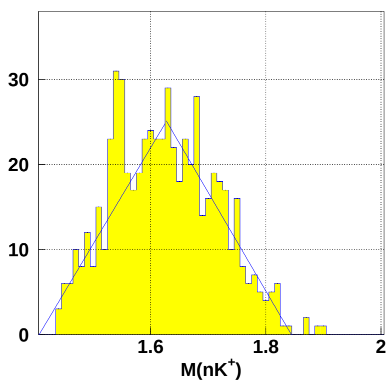

Figure 1: The K^+n invariant mass spectra as extracted from Figure 4 of the PRL. The triangular function is the background model used for the statistical study.

In order to generate a hypothetical measurement, a spectrum is generated where the background level (as an integer) is taken as the mean of a Poisson distribution. The counts in the given bin are then thrown according to a Poisson distribution whose mean is the background level in that bin.

A simple peak-finding algorithm then scans the produced spectrum and looks for three adjacent bins whose total contents are at least some threshold level above the parent background distribution. The threshold is allowed to vary from 35 to 41 counts, and corresponds to the number of counts in the Z^+ peak in the experimental spectrum.

If the background were flat, then there would be equal probability to find a fluctuation anywhere in the spectrum. However, the triangular shape makes it more likely to have a fluctuation where there are more counts. As such, one needs to be careful in defining probabilities. In this analysis, the assumption has been made that the location of the Z^+ peak needs to be *reasonably* consistent with the SPring-8 and ITEP results. To do this, I have defined a set of windows in mass. The center of the peak is required to fall within the defined windows, which are defined in Table 1. Note that the first one is the entire spectrum. If I had to choose a reasonable window, I would probably take 4, where the center of the peak is required to be between 1.515 and 1.565. Because of the binning, this would actually require the center of the three bins to be in one of the following mass bins: 1.52, 1.53, 1.54, 1.55 or 1.56.

Window	Low Mass	High Mass
1	$1.000\, GeV/c^2$	$2.000\, GeV/c^2$
2	$1.495\, GeV/c^2$	$1.585\, GeV/c^2$
3	$1.505\, GeV/c^2$	$1.575\, GeV/c^2$
4	$1.515\, GeV/c^2$	$1.565\, GeV/c^2$
5	$1.525\, GeV/c^2$	$1.555\, GeV/c^2$

Table 1: Windows used in the analysis. Note that the first corresponds to the entire spectrum, while the 2 to 5 are more reasonable for this analysis.

In this study, 15×10^6 hypothetical spectra were generated, and the number of peaks under each condition was then tabulated. The results of this are given in Table 2. The results from can be directly read off from the table, *i.e.* the chance of a peak at least 38 counts above background in window number 4 is $(38/15 \times 10^6) = 2.5 \times 10^{-6}$, or approximately 1 *in* 400000 randomly chosen background spectra will have such a peak. Whereas the probability of finding such a peak *anywhere* in the spectra is $(1851/15 \times 10^6) = 1.2 \times 10^{-4}$, or 1 *in* 8100. Finally, for completeness, the study has been repeated with somewhat larger backgrounds. I have examined the cases where every background bin has 1 additional count, and 2 additional counts. The results of these studies are shown in Tables 3 and in 4.

The last part of this study is to attempt to define a σ measure for the significance of the peak. While the numbers in the previous tables are straightforward, I suspect that there are several ways to convert them into a σ measure. The method chosen here is as follows. We assume that we are some number of σ into the tails of a Gaussian distribution, say $\alpha \times \sigma$. We then ask: "What is α such that the probability of being that far out in the tails is equal to the probability as computed from the numbers given in Tables 2, 3 and 4?". Assuming a normalized Gaussian distribution, the probability of being at $\alpha\sigma$ or larger is described by the error function complement, *erfc(x)*,

$$\text{Probability}(\alpha) = \text{erfc}(\frac{\alpha}{\sqrt{2}})\,. \quad (3)$$

Figures 2 and 3 show plots of the resulting σs for these events. The first figures shows all results except for those from window 1, (the entire spectrum). Figure 3 is for the

Threshold	Window Number				
	1	2	3	4	5
35	7372	622	365	183	84
36	4600	357	209	99	49
37	2913	229	120	54	22
38	1851	155	86	38	15
39	1166	101	53	24	11
40	725	68	35	16	9
41	438	41	22	8	4

Table 2: The number of identified peaks in each mass window (see Table 1) which are at least as large as the given threshold. These arise from 15×10^6 hypothetical spectra.

Threshold	Window Number				
	1	2	3	4	5
35	9758	946	529	290	153
36	6328	587	307	168	88
37	4086	371	190	100	49
38	2633	230	113	61	28
39	1640	136	70	34	15
40	1018	80	48	17	7
41	643	52	31	13	5

Table 3: The number of identified peaks in each mass window (see Table 1) which are at least as large as the given threshold. These arise from 15×10^6 hypothetical spectra and have a background offset of +1 count.

special case of window 4 and offset 0. Note that there are almost no measures in which we have a *so-called* 5σ peak.

Using the number of counts in the peak as something between 38 and 39, and taking the background function with no offset, I estimate that there is approximately a 1 *in* 500000 chance of the measured peak occurring purely by chance. (Results from Table 2, and window 4). If this is converted to a σ, then I estimate that it has a statistical significance of 4.75σ. If we ask what the probability of getting a peak at exactly the right place with 40 counts above background, (window 5 with zero offset), this corresponds to $(9/15 \times 10^6)$, or a 5σ effect. As a cross check for this, using a simple Poisson distribution with a mean of 45 counts. The probability of observing

Threshold	Window Number				
	1	2	3	4	5
35	13734	1509	945	564	284
36	9106	941	576	336	158
37	5824	563	329	193	86
38	3721	346	194	112	50
39	2428	214	112	58	22
40	1515	130	69	32	11
41	933	77	40	18	7

Table 4: The number of identified peaks in each mass window (see Table 1) which are at least as large as the given threshold. These arise from 15×10^6 hypothetical spectra and have a background offset of +2 count.

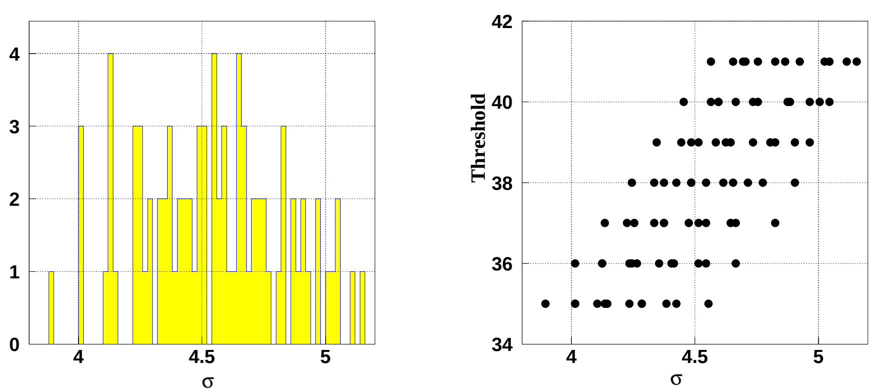

Figure 2: The left-hand figure shows the distributions of σ for all thresholds and offsets in combination with windows 2 through 5. The right-hand figure plots the thresholds versus the σ for the same set of data.

$83 = 45 + 38$ is given as:

$$P(83 : 45) = (45)^{83} e^{-45} / 83! \ . \tag{4}$$

Using the fact that $ln(n!) \approx nlnn - n$, we can estimate the chance of this fluctuation to be 1 in 366000. A number that is somewhat larger than this study indicates. The $5.4 \pm 0.6\sigma$ quoted in the PRL draft corresponds to a 1 *in* 15000000 chance of this occurring. This is somewhat of an over estimation of the peak significance, but the

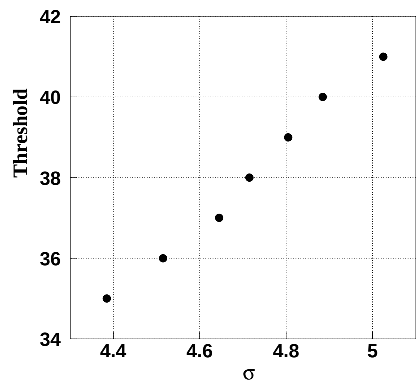

Figure 3: A plot of threshold versus σ for the case of window number 4 and and offset of 0.

method is well defined in the paper and presumably the knowledgeable reader will be able to figure this out what the tru probability is.