Statistical Analysis of Selex Double Charm Signals

Dr. Wolfgang Rolke, Dr. Angel Lopez

July 26, 2002

1 Introduction

This short note comments on the SELEX analysis as presented in their Fermilab and BEACH2002 talks. Both talks are presently available on the Web at http://fn781a.fnal.gov/selex_wac.ps.gz and at http://beach2002.physics.ubc.ca/talks/Peter_Cooper.pdf.

Here we discuss one problem with the SELEX talks, mainly the statement of the significance of the signals in terms of "number of sigmas". We believe it is only natural for the scientific community to interpret such numbers in terms of the Gaussian distribution which is the only interpretation for which these numbers have a definite correspondence to a level of significance. However, SELEX is apparently defining "sigma" as the standard deviation. Since we are dealing with Poisson statistics, the standard deviation is the square root of the number of counts. This is an exact statement. However, the probability for a deviation from the mean larger than a specified number of standard deviations is not the same for a Poisson as compared to a Gaussian distribution. It is typically much larger when the number of counts is small. Thus the appearance of bogus signals as large as those seen by SELEX is more probable than they imply by stating a number of "sigmas". The significance of the SELEX results is also lowered by the consideration that they looked for signals of any width anywhere within a large mass range.

2 General Comments

The crucial question in the Selex talks is whether or not they see a signal. From a statistical point of view this should be done by performing a hypothesis test with the null hypothesis H_o:"*There is no signal present*". Rejecting this H_o is then equivalent to claiming a discovery. The standard way to perform a hypothesis test is to compute the p-value, defined here as the probability of seeing a bump as large as or larger than the one observed in the real data when there really is no signal. The p-value will be very small for significant signals. The Selex hypothesis test is based on the quantity $S/\sqrt{B}$. The use of this variable as a measure of significance can only be justified on the basis of the Gaussian approximation to the Poisson distribution. The p-value is then found by $P(Z > S/\sqrt{B})$ where Z is a Gaussian random variable. Let us investigate how good an approximation this actually is. For this we generated 1 million observations X_i from a Poisson distribution with rate $b = 6.1$ and another 1 million observations Y_i from a Poisson with rate $\tau * b = 24.25 * 6.1 = 147.9$, Here b is the background rate in the signal region and τ is the ratio between the sideband and signal backgrounds. The numbers were chosen to coincide roughly with the most significant of the SELEX bumps. We want to see what the probability is of seeing a "signal" of a certain size when there is really only background. Such a large sample size is necessary because we are going to probe the extreme tails of this distribution. For each of these 1 million pairs we find $S/\sqrt{B} = (X_i - Y_i/\tau)/\sqrt{Y_i/\tau}$. Next we count the number of $S/\sqrt{B} > t$, were t varies from 3 to 6.5. These numbers are then compared to the exact quantiles from a Gaussian distribution, that is the probabilities $P(Z > t)$. The results are in the following table:

t	$P(Z > t)$	$\# S/\sqrt{B} > t$	Ratio $\frac{\# S/\sqrt{B} > t}{P(Z>t)}$
3.0	$1.35 * 10^{-3}$	$5.46 * 10^{-3}$	4
3.5	$2.33 * 10^{-4}$	$1.85 * 10^{-3}$	8
4.0	$3.17 * 10^{-5}$	$5.85 * 10^{-4}$	19
4.5	$3.40 * 10^{-6}$	$1.92 * 10^{-4}$	57
5.0	$2.87 * 10^{-7}$	$5.7 * 10^{-5}$	199
5.5	$1.90 * 10^{-8}$	$1.6 * 10^{-5}$	842
6.0	$9.87 * 10^{-10}$	$7 * 10^{-6}$	7095
6.5	$4.02 * 10^{-11}$	$2 * 10^{-6}$	49800

What does this mean for the significance? Take the particular case when $S/\sqrt{B} = 6.3$. From (a more detailed version of) the table above this corresponds to $p = 3 * 10^{-6}$. This in turn corresponds to a significance level of a Gaussian distribution of about 4.5σ. In such a case, we argue that SELEX should be quoting a value of 4.5σ for the significance instead of the value of 6.3.

3 Consideration of the signal position

In the BEACH2002 talk, SELEX brings up the point (worded differently) that the appropriate p-value for their "discovery" should be adjusted for the fact that they did not know the exact location where these signals had to be. They looked for signals in a wide range of invariant mass. This increases the probability of a bogus signal appearing from the background thus increasing the p-value. In that talk, SELEX says that the adjustment is accomplished by multiplying the basic p-value by the ratio between the size of the search region to the signal region. A much more complicated calculation is necessary to do the appropriate adjustment for this effect but in the analysis below we have used the SELEX prescription to obtain a quick result which reflects the effect somewhat but still severely underestimates the correct p-value. In order to do this, one needs to know the ratio of the search region to the signal region (here called r). SELEX only quotes the value (100) for their most significant bump but, with a signal region of $10MeV$, this corresponds to a search region of $1GeV$. We have assumed this is also the size of the search region for the other two bumps and have calculated the corresponding p values. In order to distinguish the p-values and sigmas obtained when this effect has been considered we shall use the adjective "total" as opposed to "basic" for the case when it has not.

4 Analysis Using Alternative Methods

Because of the duality of confidence intervals and hypothesis tests we carry out the hypothesis test by simply finding confidence intervals with a variety of confidence levels α's until we find the threshold α where the lower confidence limit changes from a positive number to 0. The p-value is then given by $p = 1-\alpha$. For a given p-value, one can easily calculate a significance in terms of σ. As an example consider the first case below. There we have $x = 22$ events in the signal region and an estimated background rate of 6.1. The Feldman-Cousins lower limit is equal to 0 for $\alpha = 0.99999937$ and greater than 0 for $\alpha = 0.9999993$. Therefore the p-value is $6.7 * 10^{-7}$ and by equivalence to confidence intervals from a Gaussian distribution this amounts to a 4.9σ effect.

We have used both the unified method of Feldman and Cousins and the method of Rolke and Lopez to calculate the significance of the Selex signals. These methods are typically used to calculate confidence limits but they can be used to calculate p-values as explained in the previous paragraph. The main advantage of these methods is that they make no prior assumption about the existence (or non-existence) of a signal. Also they both take into account the Poisson statistics correctly. The difference between them is that the Rolke-Lopez method takes into consideration the uncertainty in the background rate while Feldman-Cousins does not. For the problem at hand, the differences between them will be relatively small since the size of the sideband samples is rather large. Here the two methods give slightly different limits and thus slightly different p-values.

Although we believe the Selex estimates for the numbers of signal and background events are not quite correct, in the following we have assumed that they are, in order to emphasize that the main problem with the significance claim is in the methodology used. In order to use Rolke-Lopez, one needs to know the ratio of sideband background to signal background (τ). This ratio has been determined using the Selex linear background fits and their signal and background regions. For a given number of background events in the signal region, the Rolke-Lopez results are highly insensitive to this ratio over the range of reasonable choices for the sizes of these regions.

Selex apparently made some changes between the talk at the Fermi lab and the talk at Beach 2002. The following table has the numbers from both talks:

Talk	ccd+	ccu++	ccu*++
Fermi Lab	$\frac{15.8}{\sqrt{6.2}} = 6.4\sigma$	$\frac{8}{\sqrt{2}} = 5.6\sigma$	$\frac{27}{\sqrt{35}} = 4.5\sigma$
Beach 2002	$\frac{15.9}{\sqrt{6.1}} = 6.3\sigma$	$\frac{7.6}{\sqrt{2.6}} = 4.8\sigma$	$\frac{27.4}{\sqrt{47}} = 4.0\sigma$

In the following analysis we will use the (later) numbers from the Beach 2002 talk.

4.1 SELEX ccd+ Histogram

Signal events 15.9

Estimated background events 6.1

Signal region: 3.515-3.525

normalized linear background: $y = -24.75 + 8.16x$

$\tau = 24.25$

$r = 100$

Method	Basic p	Basic σ	Total p	Total σ
Selex claimed confidence level	$1.5 * 10^{-10}$	6.3	$1.5 * 10^{-8}$	5.54
Selex true confidence level	$3 * 10^{-6}$	4.5	$3 * 10^{-4}$	3.4
Feldman-Cousins	$5.15 * 10^{-7}$	4.89	$5.15 * 10^{-5}$	3.90
Rolke-Lopez	$1.65 * 10^{-6}$	4.65	$1.65 * 10^{-4}$	3.59

4.2 SELEX ccu++ Histogram

Signal events 7.6

Estimated background events 2.6

Signal region: 3.448-3.472

normalized linear background: $y = -78.7 + 23.8x$

$\tau = 10.7$

$r = 41.7$

Method	Basic p	Basic σ	Total p	Total σ
Selex claimed confidence level	$7.9 * 10^{-7}$	4.80	$3.3 * 10^{-5}$	3.99
Selex true confidence level	$5.24 * 10^{-4}$	3.28	$2.2 * 10^{-2}$	2.02
Feldman-Cousins	$3.8 * 10^{-4}$	3.37	$1.6 * 10^{-2}$	2.15
Rolke Lopez	$1.3 * 10^{-3}$	3.02	$5.4 * 10^{-2}$	1.61

4.3 SELEX ccu*++ Histogram

Signal events 27.4

Estimated background events 47

Signal region: 3.733-3.833

normalized linear background: $y = 5.8 - 0.78x$

$\tau = 2.56$

$r = 10$

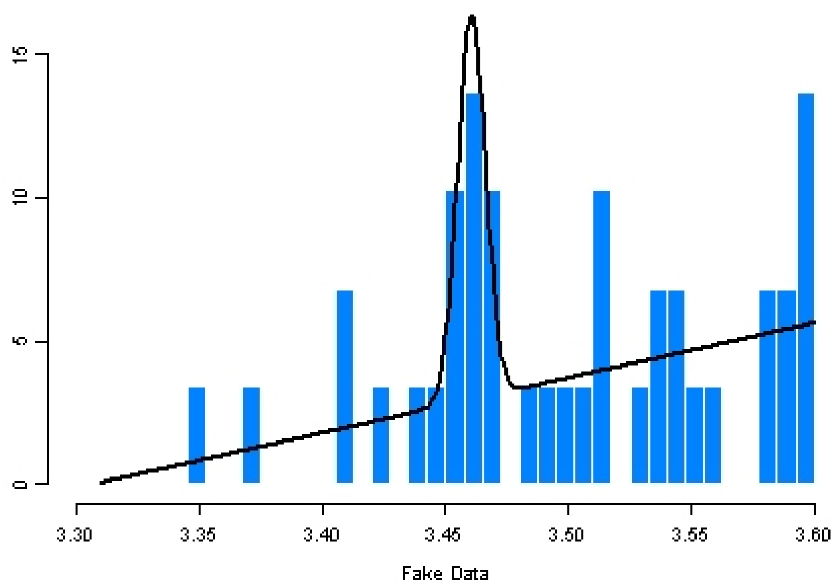
Method	Basic p	Basic σ	Total p	Total σ
Selex claimed confidence level	$3.2 * 10^{-5}$	4.0	$3.2 * 10^{-4}$	3.41
Selex true confidence level	$1.2 * 10^{-3}$	3.04	$1.2 * 10^{-2}$	2.25
Feldman-Cousins	$1.6 * 10^{-4}$	3.59	$1.6 * 10^{-3}$	2.95
Rolke Lopez	$2.5 * 10^{-3}$	2.81	$2.5 * 10^{-2}$	1.96

5 A Mini MC Study

The bumps in the Selex graphs are actually quite impressive at first glance. Is it possible to see bumps as large as these even if there is no signal at all? In order to study this question we performed the following mini MC study using the ccu++ histogram:

- Generate 39 events, as in the histogram, from the linear background density $f(x) = -78.7 + 23.8x$ on $[3.31, 3.6]$
- Check whether there are 4 consecutive bins to the left of 3.4725 with a total of events ≥ 10
- If so, check whether these 4 bins form a nice Gaussian shape
- If so, stop, otherwise start over.

We ran this MC, and on run #1742 it stopped with this:



Next we repeated this mini MC study 50 times, and we found a signal at least as nice as this one on average every 7010 runs. $1/7010 = 1.4 * 10^{-4}$ corresponds to 3.6σ. This significance level is much closer to the ones given by Feldman-Cousins and Rolke-Lopez than the one obtained from $S/\sqrt{B}$. Notice that here we used rather strong requirements for the shape as well as the size of the peak and yet found plenty of nice peaks.

If we drop the requirement of a "nice" Gaussian shape and only look for at least 11 events (9 over a background of 2), then we can find this on average every 1 in 374 runs, corresponding to 2.79 sigmas.

6 Conclusions

We arrive at the following conclusions:

- The confidence levels claimed by Selex are too large, primarily because the distribution of $S/\sqrt{B}$ in the extreme tails is not Gaussian.
- The correct way to compute the significance levels and the p-values is by using the Rolke-Lopez method. Note that computing "adjusted" p-values and significance levels through a mini MC as described in section 2 and carried out for each of the 3 signals leads to p-values very close to those given by the Rolke-Lopez method.
- As correctly pointed out by Selex, all the significance levels are too high because the signal peak was allowed to be anywhere in the mass range.
- Has Selex discovered new particles? According to our analysis their best candidate has a significance level of about 3.6σ. Whether this is sufficient for claiming a discovery is a judgment call that only the Selex group can make.