Comments on Likelihood Fits with Variable Resolution

Giovanni Punzi
Scuola Normale Superiore and INFN, 56100 Pisa, Italy

Unbinned likelihood fits are frequent in Physics, and often involve complex functions with several components. We discuss the potential pitfalls of situations where the templates used in the fit are not fixed but depend on the event observables, as it happens when the resolution of the measurement is event–dependent, and the procedure to avoid them.

When several categories of events are present in the same data sample, an unbinned Maximum Likelihood fit is often used to determine the proportion and the properties of each class of events. This procedure makes use of "templates", representing the probability distribution of the observables used in the fit for each class of events. In the simplest cases the templates are completely determined by the values assigned to the parameters of the fit, but frequently a more sophisticated approach is chosen where templates vary on an event by event basis, according to the resolution of the measurement for that particular event. These variations are due to the dependence of resolution on extra variables, that change on an event-by-event basis . This may happen, for instance, when events are recorded by a detector that has different resolutions in different regions within its acceptance.

A common example of this kind of fit in HEP is given by lifetime and/or mass fits (see [1] for a sample list of recent experimental papers), where variations in resolution occur as a consequence of different configuration of each individual decay. The same kind of issue however is likely to arise in other situations.

The purpose of this short paper is to point out some potential pitfalls in this kind of fitting procedure. I will illustrate the point with reference to a simple toy problem.

1. A TOY PROBLEM

Consider an experiment in which two types of events, A and B, can occur. Let f be the fraction of type–A events, that is, the probability of a generic event to be of type A. We want to extract a measurement of f from a given sample of data. In order to do this, we measure the value of an observable x, having the following probability distributions:

$$p(x|A) = N(0, \sigma)$$
$$p(x|B) = N(1, \sigma)$$

Where σ is a known constant and $N(\mu, \sigma)$ is the normal distribution

This problem is easily solved using an "unbinned Likelihood fit". This consists of maximizing the Likelihood function:

$$L(f) = \prod_i \left(f N(x_i, 0, \sigma) + (1-f) N(x_i, 1, \sigma) \right) \quad (1)$$

with respect to the required parameter f (here $N(x, \mu, \sigma)$ indicates the gaussian function in the variable x). This is very simple to perform with the help of a numerical maximization program.

Let's make a specific numeric example, where $f = 1/3$ and $\sigma = 1$ (see illustration in Fig. 1), and the size of the data sample is 150 events. By repeatedly generating MC samples of 150 events each, we obtain the distribution of the Maximum Likelihood estimator of f, which is shown in Fig. 2.

Its mean is 0.3368 ± 0.0041 and $SD = 0.083$, in agreement with expectations of 0.3333 and 0.088 respectively (the latter coming from Fisher information).
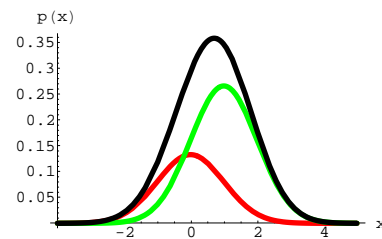

Figure 1: Probability distribution of x for the toy problem described in the text. Contribution of type–A and type–B events are also shown.
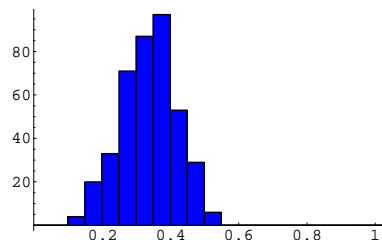

Figure 2: Distribution of ML estimate of the fraction f of type-A events (see text)

2. A TOY PROBLEM, WITH VARIABLE RESOLUTION

Let's now suppose that the resolution of x is not constant, but rather depends on the event: we are assuming that each event x_i comes together with an individual value of σ (let it be σ_i). This situation is encountered in many real–life problems, and the common approach found in the literature is to simply modify the Likelihood function as follows:

$$L(f) = \prod_i fN(x_i, 0, \sigma_i) + (1-f)N(x_i, 1, \sigma_i)) \quad (2)$$

This looks like a pretty obvious generalization of expression (1). To test it in our toy problem, we modified our toy MC from previous example, by making σ fluctuate at each event within an arbitrarily chosen range (1.0 to 3.0), and again made repeated simulated experiments of 150 events each, maximizing the Likelihood expression (2) to estimate f. The result of this test is shown in Fig. 3, and rather surprisingly, shows a very large bias with respect to the true value of f.
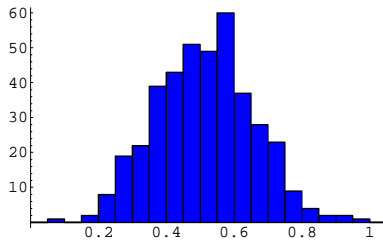

Figure 3: Distribution of ML estimate of the fraction f of type-A events, obtained from a "conditional Likelihood"

This may seem really odd, until one realizes that this new problem is very different from the previous one. Our problem now has actually two observables: each observation consists of the pair of values (x_i, σ_i) rather than just x_i, and its probability density depends on both. This means that the Likelihood must now be written based on the probability distributions of the (x_i, σ_i) *pair*:

$$L(f) = \prod_i fp(x_i, \sigma_i|A) + (1-f)p(x_i, \sigma_i|B) \quad (3)$$

Remembering that $p(x_i, \sigma_i|X) = p(x_i|\sigma_i, X)p(\sigma_i|X)$ we can write the correct expression of the Likelihood for our problem as:

$$L(f) = \prod_i fN(x_i, 0, \sigma_i)p(\sigma_i|A)$$
$$+ (1-f)N(x_i, 1, \sigma_i)p(\sigma_i|B) \quad (4)$$

where $p(\sigma_i|X)$ is the *pdf* of σ_i for events of type X, an element that was absent in eq. (2); in fact, comparing the two expressions shows that (2) is actually the *conditional* probability distribution $p(x_i|\sigma_i, f)$ (one might

call it "conditional Likelihood") rather than the full distribution $p(x_i, \sigma_i|f)$. The difference matters for fitting unless it happens that the distribution of σ_i is the same for all types of events: $p(\sigma_i|A) = p(\sigma_i|B)$. In that case, $p(\sigma_i)$ can be factored out, and the incomplete Likelihood of eq. (2) differs from the true Likelihood just by a factor independent of f, that does not affect the maximization.

In the specific MC test reported above, we simulated a resolution 1.5 times worse for events of type B than for type A, setting the σ_i distribution as flat between 1 and 2 for A-type events, and flat between 1.5 and 3 for type-B events. We intentionally avoided saying this explicitly before, in order to put the reader in the typical situation encountered in reality, where no attention is payed to the distribution of those resolutions for the different classes of events considered in the fit. It turns out from our example that this may lead to very biased results.

In summary, expression (2) simply does not work for fitting, and by a large amount: it can be said to belong to that particular class of solutions nicely defined in [2] as 'SNW solutions'.

Conversely, if we use in fitting the correct expression of the Likelihood (eq. 4) we get the result shown in fig. 4, showing a negligible bias. The resolution of the fit is also much better, as the difference in the distributions of the σ themselves gets exploited in separating the two samples; this however is a minor point in comparison with the bias issue.
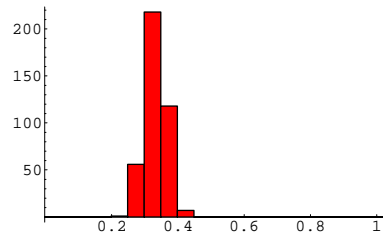

Figure 4: Distribution of ML estimate of the fraction f of type-A events, using the full Likelihood function

3. ADDITIONAL TESTS

One may wonder at what features of the distributions make for a large bias. Table I shows results for a few variants of the original problem. Tests include:

- Equal-width ranges of σ.
- Disjoint σ ranges.
- Constant, but different σ for A and B.
- Constant, and close σ's for A and B.
- Same-mean σ distribution with different widths.

Table I Results of MC fitting tests.

Resolutions		"conditional" L (2)		True Likelihood	
σ_A	σ_B	$\tilde{f}_A$	$\sigma(\tilde{f}_A)$	$\hat{f}_A$	$\sigma(\hat{f}_A)$
1.0	1.0	0.336 ± 0.003	0.08		
[1.0, 2.0]	[1.5, 3.0]	0.514 ± 0.007	0.14	0.335 ± 0.002	0.03
[1.0, 2.0]	[1.5, 2.5]	0.474 ± 0.007	0.14	0.335 ± 0.002	0.03
[1.0, 2.0]	[2.0, 3.0]	0.579 ± 0.008	0.15	0.333 ± 0.000	0.00
1.0	2.0	0.645 ± 0.006	0.12	0.333 ± 0.000	0.00
1.0	1.1	0.374 ± 0.004	0.08	0.333 ± 0.000	0.00
[0.5, 3.5]	[1.5, 2.5]	0.330 ± 0.006	0.12	0.332 ± 0.002	0.03
1.0	[1.0, 2.0]	0.482 ± 0.009	0.09	0.333 ± 0.000	0.00
(σ_A actually = 1.)		modified L (5)		True Likelihood	
1.0	[1.0, 2.0]	0.374 ± 0.004	0.08	0.333 ± 0.000	0.00
[0.5, 3.5]	[1.0, 2.0]	0.414 ± 0.004	0.08	0.332 ± 0.003	0.03

- Only one type of events has variable sigma.

In almost every tried situation we found expression (2) to return largely biased results. The exception occurs when the average σ is the same; the resolution on f is however much worse than with the correct expression. It looks like the most important element is the difference between the average values of σ for the different samples; the actual variability within each sample seems less important.

A simpler situation exists, that is pretty common in practice, where one has just one signal component over a background, and the signal distribution contains a variable sigma, while the background is represented just by a fixed template. In this case, expression (2) becomes:

$$L(f) = \prod_i fN(x_i, 0, 1) + (1 - f)N(x_i, 1, \sigma_i)) \quad (5)$$

This expression of L is of course still incorrect, but it better describes reality at least for one of the two event categories by incorporating explicitly the information that it has a fixed sigma. Here a variable template appears just in one component, and being this the simplest configuration with a variable template, it is interesting to ask whether it yields a reasonable approximation of the correct results.

If we apply this new Likelihood expression to the last tested case, ($\sigma_A = 1.0$ and $\sigma_B \in [1.0, 2.0]$), we find that the result is still biased, although to a lesser extent (Tab. I). This shows that the distribution of σ must be kept into account even in the simplest situation, where it appears in only one component of the fit.

The mechanism underlying this problem is easier to see by looking at a variant of the previous case. Suppose that resolutions are the same as above, but for events of type-A the variable σ_i is distributed over a wide range (0.5-3-5); this is not the actual value of the resolution for those events, that is still fixed at 1, so for type-A events it represents just an additional meaningless number. This is a definite possibility in a real case, where the nature of type–A events may be so different from type–B to produce meaningless values for the resolution estimator σ_i, that was designed to work for type–B events – remember that the distribution of A is given as fixed. Note that the expression used (5) does know that much, and correctly disregards the value of σ_i in the A hypothesis. For events of type B, the variable σ_i correctly represents the sigma, event by event, of the observable x, and the L function correctly accounts for this, too.

It may come as a surprise that the result is largely biased. The reason for this rather spectacular failure is that the second piece of L, related to B-type events, gets confused by the presence of the events of type–A with meaningless values of sigma: they unavoidably enter both terms of L during the calculation. The conclusion is: whenever you include σ_i in your Likelihood expression, even for just *one* class of events, you must also account for its distribution, and you must do so for *all* event classes.

4. CONCLUSIONS

Whenever the templates used in a multi-component fit depend on additional observables, one should always use the correct, complete Likelihood expression (4), including the explicit distributions of all observables for all classes of events. This is necessary even if just one of the components is based on a variable σ. The simpler expressions that are commonly used should be considered unreliable unless one can show that the distribution of the variable σ is the same for all components.

A more general consideration suggested by the examples discussed above is that one should always be wary of "intuitive" modifications of a Likelihood function. For every given problem there is only one correct expression for the Likelihood (up to a multiplicative constant factor), and it is crucial to verify in every case that the expression used is the right one, rather than rely on intuition.

References

[1] Recent examples of variable-resolution fits in HEP can be found in: B. Aubert et al. [Babar Collab.], Phys. Rev. Lett. 91 (2003) 121801. K.Abe et al. [Belle Collab.], Phys. Rev. Lett. 88 (2002) 171801. P. Abreu et al. [Delphi Collab.] , Eur.Phys. J. C16 (2000) 555. ALEPH Collab., Phys.Lett. B492 (2000) 275-287. M. Paulini, Int. J. Mod.Phys. A14 (1999) 2791-2886.

[2] J. Heinrich, these proceedings.