INTRODUCTION AND STATEMENT OF THE PROBLEM

F. James
CERN, Geneva, Switzerland

As far as I know, this is the first workshop or conference ever held by physicists and exclusively devoted to questions of statistics. But it is probably not the last, because there is already talk of a second one, in March at Fermilab.

It is always difficult to be the first ones to do something and so Louis and I are counting on you participants to help make this meeting a success. Our goal, of course, is to have everybody agree on one method for setting confidence limits, but as nobody really thinks that is possible, let us set a more modest goal of improving general understanding of this problem, so that, even if we don’t agree with what other people are doing, we can at least understand what they are doing, and maybe we can also understand our own methods a bit better when they have been confronted with real criticism. So that’s the first idea, instead of just saying what is good about your method, you should say as well what’s bad about it, and in case you don’t, everyone is encouraged to point out the bad properties in the discussion.

That leads me to the second point: This is a workshop, it’s not a conference. That means that we are all supposed to participate and we want everybody’s views. This may be a problem because there are 136 of us and so it’s not going to be easy in two days to get the views of everybody. The session chairmen will have a certain responsibility here to let everybody make comments and still keep some order.

Everything will be recorded. We have an excellent technical team here recording everything on both audio and video. This will help us in preparing proceedings. You can also help us by identifying yourself when you give a comment, so we can put all the discussion in the proceedings. We also want the speakers to provide written summaries of their talks for the proceedings. And the last session tomorrow will be devoted entirely to discussion, so everybody will be able to present his views, even those who are not official speakers. We are very pleased that all the people whom we invited have accepted, many people have come from a long distance, so this promises to be an interesting workshop. We are surprised and very pleased at the interest that everyone is showing.

Now, we need some ground rules, and one ground rule which we ought to have is that people should give the mathematical basis for their method. All methods should be based on some accepted methods of statistics. Of course, we are physicists, we’re not statisticians, and we want the emphasis to be on physics, but we do have among us at least two people who are real statisticians, and they have the job of keeping us honest, that is making sure we remain in accordance with some accepted statistical practice.

You will soon see who these two statisticians are because they are likely to use some words we physicists are not used to, but their participation is very important. We should not be in the business of inventing new statistical methods: After all, statistics have been around for a long time, some people think since Bayes, others think it’s been since Karl Pearson and R. A. Fisher, but whichever you choose, it’s been at least 100 years, and so it’s unlikely we would need to invent completely new things that nobody has ever thought of. To be more precise, we want to know whether you are using a frequentist (classical) basis for your method, or if you are building on a Bayesian foundation. Physicists have been known to confuse the two, sometimes thinking they are doing a frequentist analysis and in fact introducing Bayesian ideas. This may be acceptable, but at least they should know that they are doing it.

If your method has a frequentist (classical) basis, you should give its frequentist properties, and you know that for confidence limits, the most important one is the coverage. Then there is the well-known problem with the frequentist method that in order to be wrong 10% of the time (for 90% coverage), some methods require you occasionally to give a limit you know is wrong, and we have to know how you plan to handle that.
If you want to use Bayesian methods, you have to explain how you are going to solve the well-known Bayesian problems, one of which is the prior distribution. Do you propose a subjective prior distribution or do you think it’s possible to have an objective one? And how do you define probability? Do you define probability only as a degree of belief, or do you consider that it has a relation with long-term frequency, and there again, there are different schools of Bayesian thought. Howson and Urbach, for example make the statement [in “Bayesian Reasoning in Science”, *Nature*, 350 (1991) 371–374]: “you can not derive from the probability of an event even the approximate frequency with which that event will appear in any actual run of trials, however long”. They are of course talking about the Bayesian concept of probability, which they dissociate completely from the idea of frequency. There is a serious problem with quantum mechanics if you accept that point of view, since probability in quantum mechanics is frequentist probability, and is defined as a long-term frequency. Bayesians will have to explain how they handle that problem, and they are warned in advance.

What can we use as criteria to judge what method is acceptable? In the end, we will have to consider how the results are going to be used. I can think of three different things that I would like to do with the result of an experiment.

1. To judge the sensitivity of the experiment,
2. To combine with other results to form unbiased averages, and
3. To combine with subjective input to draw conclusions about Nature.

First of all, it is clear that in the asymptotic situation with nice Gaussian error distributions, confidence intervals are just given by ordinary standard deviations and they are fine for all three purposes. However, as we all know, when you get close to a physical limit, or when you don’t observe any events (which is a common case now), or whenever you have a very small amount of data, then the confidence limits which are good for one of these purposes may not be good for the others, so that’s where we have to say what we really want to do with these numbers. Judging the sensitivity of an experiment is a very important thing: taxpayers want to know why we spent 10 000 000 euros on an experiment that got a bigger error than someone else who only spent 5 000 000 euros. So this is important, and as you know, Feldman and Cousins found their confidence limits were not very good for judging the sensitivity, so they proposed an additional number. This is clearly a lesson to be learned, that just two numbers, an upper and lower limit, are not enough to do everything. So it could be that we have to publish more than just a small set of numbers in order to communicate all the information we have.

We are fortunate to have the participation of a representative of the Particle Data Group here. They are the biggest consumers of confidence limits in the world. If we want to sell them confidence limits, we have to sell them something that they can use. As far as I know, nobody knows how to combine confidence limits really, and so that should be an important theme here to express results so that they can easily be combined with others. I know that some of you are interested in, for example, the Bayesian way of combining results from different experiments, which is in principle very elegant, but cannot be done with the limits people publish today, including Bayesian limits.

Of course, the third thing you want to do with the result, is to use it to make your own personal judgement. For example, when somebody publishes an upper limit on a neutrino mass, you want to combine that input with your prior probability coming from theory and other experiments, to decide whether you think the neutrino mass is really zero. How should the result be published so that you know what went into it, so you can add your previous knowledge without getting also the personal feelings of the physicists who did the experiment?

That’s all I wish to say right now. As one of the convenors, I will try to be as neutral as possible and leave the expression of particular opinions up to our distinguished speakers and participants.