Statistical notes on the problem of experimental observations near an unphysical region

F. James*

CERN, CH-1211 Geneva, Switzerland

M. Roos†

Department of High Energy Physics, University of Helsinki, SF-00170 Helsinki, Finland

(Received 26 November 1990)

We discuss the statistical problem of determining confidence intervals for a random variable from data near or within an unphysical region. As a typical example we use the case of the upper limit of the electron-neutrino mass as observed in tritium β decay. We argue that it is important to publish the value of the variable actually observed (the neutrino mass squared) for which the resolution function is known (e.g., Gaussian), even if a large part of this distribution lies in an unphysical region (negative squared mass). When having to choose between quoting a result which is unbiased and one which is physical, we argue that the unbiased result is to be preferred, largely because only unbiased results can be combined meaningfully. When a measured value falls in or near a nonphysical region, then the calculation of confidence limits is necessarily a subjective procedure which represents the physicist's personal interpretation of his results, and in that sense is less fundamental than the actual measured value.

The experimental determination of the value of a physical parameter sometimes yields an unphysical result. A well-known example is the mass m of the electron neutrino in tritium β decay. Here a function of m^2 is fitted to data points representing the electron momentum spectrum near its end point.

From such fits the random variable m^2 appears Gaussian distributed, often [1–6] with an appreciable probability in the unphysical region $m^2 < 0$. The purpose of this Brief Report is to contribute to the ongoing discussion on how to report such a result (for a recent contribution to this discussion, see Robertson and Knapp [7]). Our main motivation for reopening this discussion is the need for a generally agreed procedure to report results such that independent experimental results can be combined in a straightforward and meaningful way.

Let us first note that the experimental data may be used to answer two entirely different questions, according to the desired properties of the answer.

Objective (1): Transmit unbiased information. One desires to communicate the maximum information about a parameter value, using only the data at hand. Results obtained with this objective should have the property that they can be combined with results of other experiments in a simple way. This may be called the *classical* or *information theory* approach [8].

Objective (2): Summarize all knowledge. One desires to combine information from this experiment with theoretical knowledge or belief about the value of the parameter. In particular, one explicitly excludes unphysical regions. This may be called the *Bayesian* or *decision theory* approach [8].

It must be stressed that neither approach is wrong. For many experiments it is possible to obtain the desirable properties of both approaches, but for parameter

values near a boundary of the physical region it is easy to see that *inclusion of information about the physical region biases the result away from unphysical values*.

When it is not possible to obtain the properties of both approaches, we propose that approach (1) be employed and (2) abandoned, thus favoring the classical approach over the Bayesian. The reason is the following.

Approach (2) is appropriate when making a *decision* which depends on the parameter to be determined. Physicists make decisions, for example, when they design detectors and accelerators to do experiments with. It is natural that such a decision on a design includes all knowledge available to the physicist, including his belief about the value of the parameter to be measured. He optimizes his apparatus to be able to record those values which he believes are most likely, and this belief is necessarily subjective.

On the other hand, when a physicist publishes his results he is not making a decision, but is rather transmitting information. His colleagues expect this to be independent of any subjective beliefs, and we argue that it should even be as much as possible independent of physical constraints. The main reason is that only then can the results be unbiased and therefore easily combined with other results.

In the particular case of determining the neutrino mass, if it indeed happens that the mass is zero, then in the unbiased approach one would necessarily expect the unphysical result $m^2 < 0$ half of the time. How should one then make statistical inferences from an observation reported as a squared mass $m^2 \pm \Delta m^2$ when m^2 is negative or near the unphysical region (near in terms of Δm^2)?

It is of course always possible to choose a classical confidence level p for the squared mass such that the corresponding classical confidence limit $m_{p,cl}^2$ is in the physi-

cal region. When m^2 has a Gaussian distribution the upper $p\%$ confidence limit is

$$m_{p,\text{cl}}^2 = m^2 + Z_p \Delta m^2, \quad (1)$$

where $Z_{90} = 1.282$ and $Z_{95} = 1.645$, as is certainly well known to our readers. Information from several individual Gaussian-distributed observations of the random variable m^2 can be combined in the well-known way (from the individual measurements, not from confidence limits) as the weighted mean of the individual m^2 values, and the classical confidence limit of the mean is obtained by the same procedure as for an individual experiment.

The classical confidence limit satisfies the probability statement

$$P(m_0^2 < m_{p,\text{cl}}^2) = p. \quad (2)$$

Note that the upper limit $m_{p,\text{cl}}^2$ is a random variable because it is a function of the random data. The above statement in fact says nothing about the true value m_0^2 which is a fixed, unknown number and not a random variable. The statement actually says that $m_{p,\text{cl}}^2$ has the probability p of being larger than the true value, whatever m_0^2 really is. Another way of phrasing this is the following: if one repeated the experiment many times, always recalculating the value of $m_{p,\text{cl}}^2$, then it would be true $p\%$ of the time that $m_{p,\text{cl}}^2$ was greater than the true value m_0^2 .

Equation (2) appears to be, and is sometimes incorrectly interpreted as, a probability statement about m_0^2 , which it is not. The incorrect interpretation clearly breaks down when $m_{p,\text{cl}}^2 < 0$, whereas the strict interpretation above is still meaningful even when the upper limit is unphysical. This should in fact happen with probability $1-p$ if the true value were zero.

In the Bayesian approach one obtains a confidence limit by the following procedure. The (Gaussian) likelihood function, representing only the experimental data, is multiplied by a *prior belief* function, which is of course zero in the unphysical region and nonzero in the allowed region. The exact shape of the prior belief function in the physical region can be chosen freely, and is often taken to

be constant even though this is not a proper probability density. The product of the two functions is known as the *posterior density* of m^2 and an upper limit $m_{p,B}^2$ may be determined such that the integral under the posterior density from zero to $m_{p,B}^2$ is p . (This is the procedure suggested by the Particle Data Group [9], but it is not identified by them as being a Bayesian procedure.) Thus only the $m^2 > 0$ tail of the experimental distribution will be used. Note that one then relies heavily on the assumption of the shape of the tail, e.g., Gaussian, whereas the true shape may not be reliable further out than one or two experimental standard deviations.

When the Bayesian prior density is taken as constant in the physical region, it is always true that

$$m_{p,\text{cl}}^2 < m_{p,B}^2. \quad (3)$$

Therefore, it is often claimed [9] that $m_{p,B}^2$ is conservative. However, since the Bayesian method always contains an element of arbitrariness, one cannot quantify that conservatism in terms of a confidence level.

Turning now to the linear mass m , we encounter further problems: the probability density function of m cannot be obtained by changing variables in the standard way [8] from the known probability density function of m^2 , e.g., Gaussian, because of the presence of the unphysical region. It is common practice to quote the square root of either $m_{p,\text{cl}}^2$ or $m_{p,B}^2$ as the upper confidence limit on m , which is reasonable since confidence limits are invariant. But such a method cannot be considered quantitatively "correct" since the confidence limits in this case are arbitrary anyway.

The argument is sometimes raised that if everybody agrees on a given procedure, albeit of unknown confidence, then the procedure is no longer arbitrary and different experiments can be compared. This argument is fallacious because in the case at hand it is well known that a low confidence limit can be either the result of a precise measurement falling in the physical region, or a less precise one which happens to fall in the unphysical region.

In Table I we reproduce seven recent values of

TABLE I. Electron-neutrino mass.

Reference	m_ν^2 (eV ²)	m_ν^2 errors (eV ²)		Quoted upper 95% m_ν limit (eV)
		Statistical	Systematic	
[1]	-11	63	178 ^a	18 ^a
[2]	-57	453	118	26.8 ^a
[3]	+223	244	269 ^a	32 ^a
[4]	-65	85	65	13 ^c
[5]	-147	55	58	9.4 ^c
[6]	-158	150	103	15.4 ^c
[10]	+675	36	b	40 ^c
[1-6]	-103	56 ^d		

^aWe use the systematic errors as evaluated by the Particle Data Group [9]. For the definition of the published upper limits, see also Ref. [9].

^bSystematic error not known to us.

^cDefinition not known to us.

^dStatistical and systematic errors added in quadrature.

$m^2 \pm \Delta m^2$ for the electron neutrino, obtained from the end point of the tritium β -decay spectrum. For illustration, Table I also gives the quoted “upper 95% confidence limits” for the linear mass, usually derived by some sort of Bayesian procedure, different for each experiment, and certainly of unknown confidence.

As can be seen, the classical weighted mean of the six m^2 results for which the systematic error is known [1–6] is perfectly reasonable, and does not indicate any serious discrepancy among them. It is another problem, and not a statistical one, to understand why the weighted mean comes out so strongly negative. Of course, if all seven measurements could be averaged correctly using both statistical and systematic errors, the result would certainly be closer to zero or even positive, in view of the one large positive value for which we do not know the systematic error [10].

In conclusion, we recommend the following.

(1) Experimentalists should always publish the measured value and error (both statistical and systematic) of

whatever parameter is approximately Gaussian distributed, even if the value is unphysical. For the neutrino mass as measured in tritium beta decay, the appropriate parameter is m^2 .

(2) If the measured value is consistent with zero, then optionally the physicist may report an upper confidence limit, either classical or Bayesian according to his taste.

If recommendation (1) is followed, then the reader of the publication can use the published values to calculate the confidence limit himself, using whatever method suits *his* taste. He can also combine the published values with those of other experiments, and calculate an upper limit for the combined data, again using the method of his choice.

If, on the other hand, only an upper limit is published, the result cannot be combined with those of other experiments in a straightforward way; moreover, since the method of calculating the upper limit is not unique, it may be difficult to interpret the results correctly or to assess the true accuracy of the experiment.

*Electronic address: james@cernapo.cern.ch INTERNET.

†Electronic address: roos@phcu.helsinki.fi. INTERNET.

[1] M. Fritschi *et al.*, Phys. Lett. B **173**, 485 (1986).

[2] J. F. Wilkerson *et al.*, Phys. Rev. Lett. **58**, 2023 (1987).

[3] H. Kawakami *et al.*, J. Phys. Soc. Jpn. **57**, 2873 (1988).

[4] H. Kawakami *et al.*, Phys. Lett. B **256**, 105 (1991).

[5] T. J. Bowles *et al.*, Los Alamos report presented by J. F. Wilkerson at the Neutrino '90 Conference, Geneva, 1990; in *Proceedings of the 25th International Conference on High Energy Physics*, Singapore, 1990, edited by K. K. Phua and Y. Yamaguchi (World Scientific, Singapore,

1991).

[6] E. Holzschuh, presented at Neutrino '90 Conference, Geneva, 1990 (unpublished).

[7] R. G. H. Robertson and D. A. Knapp, Annu. Rev. Nucl. Part. Sci. **38**, 185 (1988).

[8] W. T. Eadie, D. Drijard, F. E. James, M. Roos, and B. Sadoulet, *Statistical Methods in Experimental Physics* (North-Holland, Amsterdam, 1971).

[9] Particle Data Group, J. J. Hernández *et al.*, Phys. Lett. B **239** 1 (1990).

[10] S. Boris *et al.*, Phys. Rev. Lett. **58**, 2019 (1987).