14 Dec 2016

Standard Deviation for PDG Mass Data

M. J. Germuska
Retired, 47 Clifden Road, Worminghall, HP18 9JR, UK.
mgermuska@aol.com, phone: +(44) 1844 339754

Abstract

This paper analyses the data for the masses of elementary particles provided by the Particles Data Group (PDG). It finds evidence that the best mass estimates are not based solely on statistics but also on overall consistency, that sometimes results in skewed minimum and maximum mass limits. The paper also points out to some other quirks that result in minimum and maximum mass limits which are far from the statistical standard deviation. A statistical method is proposed to compute the standard deviation in such cases and when PDG does not provide any limits.

Keywords: PDG, elementary particles, mass estimates, statistics, standard deviation, standard error

1. Introduction

A recent paper by the author presented a new theory of elementary particles [1] that provides a formula for the relationship between spin and mass. The formula may be tested using the existing mass data provided by the Particle Data Group (PDG) [2], but there are idiosyncrasies and a significant number of particles are not assigned min and max limits. This paper provides a method for estimating the standard deviation in cases where PDG estimates are missing or for some reason are unsuitable.

Each particle for which credible data have been published has its own section in PDG "Particle Listings". Particles for which more accurate information has been compiled, usually including: "best mass estimate", "minimum mass", "maximum mass", "spin" and "parity" appear also in the "Particles Summary". The mass estimates are based on the published measurements included in the Listings.

The estimates are not derived by purely statistical means, but also other considerations such as the overall consistency. For example, the values for Σ particles take into consideration the values of Λ particles. The result of such adjustments sometimes the best mass is not be the average of the min and max limits, hence the limits are not the standard deviation. In some extreme cases a mass limit coincides with the best mass, as for example in the case of Λ(1830) where the best mass = max mass.

When the limits are skewed it causes problems with the "relative prediction error" (RPE), i.e. the ratio of the absolute prediction error divided by PDG + or – error as required. In the example given above this leads to an infinite RPE. In cases where the mass limits are less skewed, RPE is finite but may be equally unacceptable. In such cases we wish to compute the standard deviation using the published measurements appearing in the PDG Listings. We may also wish to do so when there are no published PDG mass limits, which applies to a vast majority of the particles not included in PDG Summary.

Each author usually provides the best mass estimate, the min and max limits and in a minority of cases the number of events on which the estimates are based. It appears that for the particles where each author provides the number of events PDG sometimes uses the standard error instead of the standard deviation, but not always. When the standard error is used then for consistency we wish to compute the standard deviation, ignoring the number of events observed by individual authors, because the standard error is smaller than the standard deviation by the factor $1/\sqrt{N}$, where N is the size of the sample.

The mass measurements for a given particle use different reactions, some authors use better equipment than others and taking all this into consideration it will be assumed here that the most unbiased method is to take all measurements at their face value ignoring all factors mentioned above.

2. Probability and statistics

Our objective in this paper is to use the mass estimates for a given elementary particle published by different authors that are included in the *PDG Listings* to obtain the collective estimate of the mass mean and the standard deviation.

We start with the basic principles of probability and statistics [3]. The mean and variance of continuous probability distribution are given by:

$$\text{mean} = \mu = \int x \cdot p(x)dx \tag{1.1}$$

$$\text{variance} = \sigma^2 = \int (x-\mu)^2 \cdot p(x)dx \tag{1.2}$$

$$\text{standard deviation} = \sigma \tag{1.3}$$

We can use the formulas above to derive μ and σ for different probability distributions, e.g. Normal, Beta, Gamma, Student's t, etc. Some basic properties of μ and σ for two independent variables x and y involving constants *a*, *b* are

$$\mu(ax+by) = a\mu(x) + b\mu(y) \tag{1.4}$$

$$\sigma^2(ax+by) = a^2\sigma^2(x) + b^2\sigma^2(y) \tag{1.5}$$

If we have a sample of *n* values x_1, x_2, … x_n then the best unbiased estimators of μ and σ are

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1.6}$$

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu)^2 \tag{1.7}$$

There is frequently some confusion between the above estimator of the population variance and the quantity known as the *Sample Variance* which is usually denoted by s^2 and is defined

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 \tag{1.8}$$

This is used in some statistical procedures and the distinction between the two is important. The *Sample Variance* is a <u>biased estimator</u> of the population variance and we shall not use it this paper.

3. Weighted method

The articles that publish results of some experiments aimed at determining the mass of a particle give the best mass estimate *m* together with the mass limits, i.e. minimum and maximum. Often in addition or instead of the mass limits the authors quote the errors δ^- and δ^+. If for a given particle we have *k* published references then we wish to find the collective symmetrical limits $\mu \pm \sigma$ that may be considered to define the normal distribution.

One way of estimating the collective mean and variance is the method of weighted mean and variance as described in a PDG report [4] which implies that PDG use this method, if not always then at least in some circumstances. If the m_i have different known variances σ_i^2 then the weighted mean is computed using the formula

$$\mu(m) = \frac{1}{w}\sum_{i=1}^{n} w_i m_i \tag{2.1}$$

where $$w_i = \frac{1}{\sigma_i^2} \qquad \text{and} \qquad w = \sum_{i=1}^{n} w_i \tag{2.2}$$

The standard deviation σ is

$$\sigma = \frac{1}{\sqrt{w}} \tag{2.3}$$

This method is not suitable for our needs mainly because of the small number of measurements that are usually available for a particle, which is mostly less than half a dozen. This will be demonstrated on a particular example, i.e. elementary particle L(2350). The PDG listing for this particle provides three mass measurements with the following values in MeV:

2370±50
2365±20
2358± 6

These measurements result in the following weighted μ and σ

$\mu = 2359$ $\qquad$ $\sigma = 6$

We note that $\mu = 2359$ is almost exactly the same as the best mass estimate m_3 =2358 and $\sigma = 6$ is exactly the same δ_3! Thus the first two measurements with greater errors δ have been completely ignored! On these grounds we conclude that this method is not suitable for our needs because it is not unusual for particles to have only three or a handful of measurements.

4. Twin variance method

Another way of finding μ is to use the un-weighted mean of the best mass estimates *m*. However, for σ we need to take into account two factors, the spread of the best mass estimates *m* and the spread of the Δ values $(\min_i - \mu)$ and $(\max_i - \mu)$. Thus we have

$$\mu = \mu(m) \qquad \text{and} \qquad \sigma^2 = \sigma(m)^2 + \sigma(\Delta)^2 \tag{3.1}$$

where

$$\mu(m) = \frac{1}{k}\sum_{j=1}^{k} m_j \tag{3.2}$$

$$\sigma(m)^2 = \frac{1}{k-1}\sum_{j=1}^{k}(m_j - \mu(m))^2 \tag{3.3}$$

and

$$\Delta_i = \min{}_i - \mu(m) \qquad \text{for } i = 1, 2 \ldots k \tag{3.4a}$$

$$\Delta_i = \max{}_i - \mu(m) \qquad \text{for } i = k+1, \ldots 2k \tag{3.4b}$$

In order to find σ(Δ) we first need to find μ(Δ)

$$\mu(\Delta) = \frac{1}{2k}\sum_{i=1}^{2k}\Delta_i \tag{3.5a}$$

$$\mu(\Delta) = \frac{1}{2k}\left(\sum_{i=1}^{k}(\min{}_i - \mu(m)) + \sum_{j=1}^{k}(\max{}_j - \mu(m))\right) \quad \text{where} \quad j = i - k \tag{3.5b}$$

$$\sigma(\Delta) = \frac{1}{2k-2}\sum_{i=1}^{2k}(\Delta_i - \mu(\Delta))^2 \tag{3.6a}$$

$$\sigma(\Delta)^2 = \frac{1}{2k-2}\left(\sum_{i=1}^{k}(\min{}_i - \mu(m) - \mu(\Delta))^2 + \sum_{j=1}^{k}(\max{}_j - \mu(m) - \mu(\Delta))^2\right) \tag{3.6b}$$

In the PDG data reported by individual authors it is an exception when for a particle measurement $|\delta^-| \neq \delta^+$. Thus $|\min_i - \mu(m)|$ is close to the corresponding $(\max_j - \mu(m))$, i.e. approximately δ_i. Therefore there is one less degree of freedom and to reflect this the divisor $2k - 2$ is used above instead of $2k - 1$.

For a small number of authors, for example N(2060) with three mass measurements, we obtain σ which leads to a smaller min mass than the smallest min mass in the data, and a greater max mass than the largest max mass in the data. Thus this method cannot be right.

The reason why this method is wrong is that formula (1.5) for adding two variances applies strictly only to very large samples. When the sample is small the formula is grossly inaccurate, as can be easily verified using two random variables with three values each.

5. Approximate formulas

From the results in the previous section we conclude that unless we have a very large number of authors contributing to a given particle we must drop from the formula (3.1) for σ^2 either $\sigma(m)^2$ or $\sigma(\Delta)^2$.

Using the PDG data there is a consistent pattern, namely that $\sigma(m)^2$ is usually smaller than $\sigma(\Delta)^2$ by an order of magnitude and occasionally even smaller. Thus, if we were to use σ(m) we would obtain min and max limits similar to the weighted method. But this also means that dropping σ(m) will usually not make much difference. Thus we need to re-examine the formula for σ(Δ).

Let us therefore re-write formula (3.5b) for μ(Δ) in a way that is easier to understand, namely

$$\mu(\Delta) = \mu\left(\frac{\min + \max}{2}\right) - \mu(m) \qquad (4.1)$$

We see that if for each published reference we have $|\delta^-| = \delta^+$ then we have the best mass estimates values m = (min + max)/2 and substituting this into (4.1) above we get

$$\mu(\Delta) = 0 \qquad (4.2)$$

There is a very small percentage of cases when this condition is not satisfied, hence μ(Δ) is always approximately zero. Substituting the above to formula (3.6b) we find

$$\sigma(\Delta)^2 \cong \frac{1}{2k-2}\left(\sum_{i=1}^{k}(\min{}_i - \mu(m))^2 + \sum_{j=1}^{k}(\max{}_j - \mu(m))^2\right) \qquad (4.3)$$

If for a given measurement i the best mass m_i does not coincide with the mean mass $\mu(m)$ then we have $(\min_i - \mu(m))^2$ different from $(\max_i - \mu(m))^2$ and their sum is greater than the sum of $(\min_i - m_i)^2$ and $(\max_i - m_i)^2$ since as a rule they are equal. Thus the formula giving the <u>minimum possible</u> σ(Δ) is

$$\sigma^2 = \sigma(\Delta)^2 = \frac{1}{k-1}\left(\sum_{i=1}^{k}\delta_i^{\,2}\right) \qquad (4.4)$$

There are occasions when most authors over-estimate the accuracy of their own experiments and as a result disagree among themselves on the best mass estimate. This results in an unrealistic σ(Δ) that is much smaller than σ(m). On such occasions we reserve the freedom to use σ(m)

$$\sigma^2 = \sigma(m)^2 = \frac{1}{k-1}\sum_{j=1}^{k}(m_j - \mu(m))^2 \qquad (4.5)$$

In conclusion, if it is necessary to calculate the standard deviation for PDG mass data we will use formulas (4.4) or (4.5) whichever give greater σ as otherwise we may obtain similar results to the weighted method, which is unacceptable. In practice it means that we will mostly use (4.4) and only very occasionally (4.5) .

Acknowledgements

I would like to thank my wife Jo for her loving care without which this work would not have been possible. I thank God for everything.

References

[1] Germuska M J 2016 Vir Theory of Elementary Particles *viXra:1612.0236 [High Energy Particle Physics]*

[2] Particles Data Group (PDG) 2015 http://www-pdg.lbl.gov/

[3] Cooper B E 1969 *Statistics for Experimentalists* (Pergamon Press) p.29, 36, 46, 47, 50

[4] Cowan G 2015 http://pdg.lbl.gov/2015/reviews/rpp2015-rev-statistics.pdf p.4