The standard deviation of the mean

These notes provide some clarification on the distinction between the standard deviation and the standard deviation of the mean.

1. The sample mean and variance

Consider a random variable x and the corresponding probability distribution $p(x)$. For convenience, we consider the case of a discrete random variable, although the generalization to continuous random variables is straightforward. Given $p(x)$, one can easily compute the expectation value and the variance,

$$E(x) \equiv \mu = \sum_x x p(x)\,, \tag{1}$$

$$\text{Var}(x) \equiv \sigma^2 = \sum_x (x-\mu)^2 p(x) = E(x^2) - [E(x)]^2\,. \tag{2}$$

The standard deviation of x is denoted by $\sigma \equiv \sqrt{\text{Var}(x)}$.

In the real world, $p(x)$ is usually unknown, in which case μ and σ are unknown. However, one can perform experiments to "measure" x. Suppose n measurements are made, and the values x_1, $x_2, \ldots x_n$ are obtained. Ideally, we would like to reconstruct the probability distribution $p(x)$ from the data, but here we are interested in determining the expectation value μ and the standard deviation σ from the experimental results.

We can regard x_1, $x_2, \ldots x_n$ as independent and identically distributed random variables (often abbreviated as iid or IID random variables). These are independent, since separate measurements of x are independent of each other. These are identically distributed, since the experiment is measuring the same random variable x each time (although, of course, the outcome of each measurement will not be the same). This means that

$$E(x_i) = \mu \quad \text{and} \quad \text{Var}(x_i) = \sigma^2\,, \qquad \text{for } i = 1, 2, 3, \ldots, n\,.$$

Of course, the above information is not very practical, since a priori we do not know the values of μ and σ.

Having made n independent measurements, we would like to ascertain the best possible estimates for μ and σ. In class, we defined the sample average $\overline{x}$ and the sample variance Σ^2 by

$$\overline{x} \equiv \frac{1}{n}\sum_{i=1}^n x_i\,, \tag{3}$$

$$\Sigma^2 \equiv \frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x})^2\,. \tag{4}$$

These quantities are easily computed from the data. We now assert that the sample average $\overline{x}$ provides a best estimate for the actual mean μ and the sample variance Σ^2 provides a best estimate for the actual variance σ^2. In the mathematical statistics literature, there is some debate as to the meaning of the word "best." In the present context, the word "best" simply means that the estimates are *unbiased*, that is

$$E(\overline{x}) = \mu \quad \text{and} \quad E(\Sigma^2) = \sigma^2\,, \tag{5}$$

where the expectation values are computed assuming for a moment that we do know the underlying probability distribution $p(x)$. Let us verify eq. (5) explicitly. First, recalling that $E(cx) = cE(x)$ and $E(x+y) = E(x) + E(y)$, we have

$$E(\overline{x}) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n}\cdot n\mu = \mu\,.$$

In the Appendix, we demonstrate that $E(\Sigma^2) = \sigma^2$.

To reiterate, $\overline{x}$ provides a best estimate of the unknown μ, which is the expectation value of the random variable x. Similarly, Σ^2 provides a "best" estimate of the unknown σ^2, which is the expectation value of $(x-\mu)^2$.

2. The standard deviation of the mean

Although $\overline{x}$ provides a best estimate of the unknown μ, its determination does not tell us how likely it is that the measured value $\overline{x}$ is close to μ. After all, if I perform additional measurements of x, I would expect the value of the average $\overline{x}$ to change (although the change is expected to be small once n is large enough). Thus, what we would really like to know is the probability distribution of the random variable $\overline{x}$. Of course, since we do not know in general the expectation value and variance of x, we also do not know in general the expectation value and variance of $\overline{x}$. Indeed, we have already seen that $E(\overline{x}) = \mu$, which we do not know. Likewise, we can compute $\mathrm{Var}(\overline{x})$ as follows:

$$\mathrm{Var}(\overline{x}) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(x_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2}\,n\sigma^2 = \frac{\sigma^2}{n}\,, \tag{6}$$

which depends on the unknown σ^2.

However, we do have a "best" estimate for σ^2 based on our data, namely Σ^2 defined in eq. (4). Hence, we shall define the *standard deviation of the mean* (also called the *standard error*) to be σ_m, where

$$\sigma_m^2 \equiv \frac{\Sigma^2}{n} = \frac{1}{n(n-1)}\sum_{i=1}^{n}(x_i - \overline{x})^2\,. \tag{7}$$

The experimentalist now concludes after taking data and obtaining the values $x_1, x_2, \ldots, x_n$ after n measurements, that the best estimate of the mean is

$$\overline{x} \pm \sigma_m\,.$$

If there is a theoretical value of μ to compare this to, the experimentalist can now make statements involving confidence intervals (e.g., the probability that the data is consistent with the theoretical expectation), as discussed in Boas.

It is very important to distinguish σ_m, which is obtained from data and σ which is the unknown standard deviation of the random variable x. We have $\sigma^2 = \mathrm{Var}(x)$, which is determined by the probability distribution $p(x)$ and does not depend on the number of measurements performed by the experimentalist. The experimentalist can make an estimate for σ^2, namely Σ^2 given by eq. (4). It may look like Σ^2 depends on n, but the dependence is pretty weak (if n is large). After all, $\overline{x}$ also depends weakly on n (if n is large), which provides the best estimate for μ. However, $\sigma_m^2 = \Sigma^2/n$ depends strongly on n. The more measurements that are made, the smaller σ_m^2 is. This is not surprising, since one expects that the larger n is, the better $\overline{x}$ is as an estimate for μ. As emphasized above, σ_m^2 is an estimate of the variance of $\overline{x}$, which is obviously not the same as the variance of x [they differ by a factor of n as shown in eq. (6)]. Equivalently, Σ is an estimate of the uncertainty in a single measurement of the random variable x, whereas σ_m is an estimate on the uncertainty of the mean value of the random variable x as determined by n measurements.

A simple example illustrates the above discussion. Suppose that $p(x)$ is the binomial distribution with probability p that a tossed coin will land on heads. Define the random variable,

$$x = \begin{cases} 1\,, & \text{the coin lands on heads}\,, \\ 0\,, & \text{the coin lands on tails}\,. \end{cases}$$

Given this coin, the experimentalist is asked to determine the mean $\mu = p$ and the variance $\sigma^2 = p(1-p)$ by flipping the coin n times. After n flips, the experimentalist obtains a data set, $x_1,\ x_2, \ldots, x_n$, which is a series of 1s and 0s. From this data, the experimentalist computes $\overline{x}$ which is equal to the number of heads divided by n. The experimentalist also computes Σ using eq. (4) and σ_m using eq. (7). The experimentalist concludes that the probability p of the coin (i.e, the true mean μ) is $\overline{x} \pm \sigma_m$, where the error bars represent a 68% confidence interval, corresponding to a one standard deviation of the mean uncertainty. Clearly, the large n is (i.e. more coin flips), the smaller the corresponding standard error σ_m, and consequently the more reliable $\overline{x}$ is as an estimate of the probability p of the coin. Likewise, the best estimate for σ^2 is given by Σ^2. By the way, the latter determination also has an error associated with it, which I briefly discuss in Section 3 of these notes.

References 1 and 2 provide a cogent discussion of the differences between standard deviation and the standard deviation of the mean. In particular, reference 1 is a superb treatment of error analysis written specifically for physicists at an elementary level.

3. The standard deviation of the variance

Although Σ^2 provides a “best” estimate of the unknown σ^2, this does not tell us how likely it is that the measured value Σ^2 is close to σ^2. After all, if I perform additional measurements of x, I would expect the value of the average Σ^2 to change (although the change is expected to be small once n is large enough). Thus, what we would really like to know is the probability distribution of the random variable Σ^2. Of course, since we do not know in general the expectation value and variance of x, we also do not know in general the expectation value and variance of Σ^2. Indeed, we have already seen that $E(\overline{\Sigma^2}) = \sigma^2$, which we do not know. Likewise, one can compute $\mathrm{Var}(\Sigma^2)$, which depends in general on σ^2 and on $E(x^4)$. However, it may be of some interest to consider the case of a normal distribution, since the central limit theorem can be applied if n is large enough. In this case, it is a straightforward exercise to show that $E(x^4) = 3\sigma^4$, in which case $\mathrm{Var}(\Sigma^2)$ depends only on σ. The result (obtained in Appendix E of reference 1 and Appendix C of reference 3) is:

$$\mathrm{Var}(\Sigma^2) = \frac{\sigma^2}{2(n-1)}\,,$$

which again depends on the unknown σ^2. However, we can again employ the “best” estimate for σ^2 based on our data, namely Σ^2. Thus, we conclude that under the assumption that $p(x)$ is the normal distribution of unknown mean and variance, then the “best” estimate of the standard deviation of the variance of the random variable x obtained from our data is given by σ_v, where

$$\sigma_v^2 = \frac{\Sigma^2}{2(n-1)} = \frac{1}{2(n-1)^2}\sum_{i=1}^{n}(x_i - \overline{x})^2\,.$$

As in the case of σ_m, we see that σ_v also can be reduced in size by performing more measurements (i.e. by taking n larger). However, in practice σ_v (sometimes called the “error of the error”) is not often employed in experimental analyses.

References

1. John R. Taylor, *An Introduction to Error Analysis: the study of uncertainties in physical measurements*, 2nd edition (University Science Books, Sausalito, CA, 1997).

2. David L. Streiner, *Maintaining Standards: Differences between the Standard Deviation and the Standard Error, and When to Use Each*, Canadian Journal of Psychiatry, **41** (1996) pp. 498–502.

3. Jörg W. Müller, *Some Second Thoughts on Error Statements*, Nuclear Instruments and Methods **163** (1979) 241–251.

APPENDIX: Proof that $\boldsymbol{E(\Sigma^2) = \sigma^2}$

Starting with eq. (4), we shall compute

$$E(\Sigma^2) = \frac{1}{n-1}\sum_{i=1}^{n} E[(x_i - \overline{x})^2]\,. \tag{8}$$

It is convenient to rewrite the above equation by noting that

$$\mathrm{Var}(x_i - \overline{x}) = E[(x_i - \overline{x})^2] - [E(x_i - \overline{x})]^2 = E[(x_i - \overline{x})^2]\,,$$

after using

$$E(x_i - \overline{x}) = E(x_i) - E(\overline{x}) = \mu - \mu = 0\,.$$

Thus, eq. (8) can be rewritten as

$$E(\Sigma^2) = \frac{1}{n-1}\sum_{i=1}^{n} \mathrm{Var}(x_i - \overline{x})\,.$$

To evaluate the above expression, we shall use [1]

$$\mathrm{Var}(cx) = c^2\,\mathrm{Var}(x)\,, \qquad \mathrm{Var}(x+y) = \mathrm{Var}(x) + \mathrm{Var}(y)\,, \tag{9}$$

where the latter holds under the assumption that x and y are independent random variables. Since x_i and $\overline{x}$ are not independent random variables (since $\overline{x}$ contains x_i in its definition), we must perform the following manipulation,

$$x_i - \overline{x} = x_i - \frac{1}{n}\sum_{i=1}^{n} x_i = \left(\frac{n-1}{n}\right) x_i - \frac{1}{n}\sum_{j\neq i} x_j\,.$$

Consequently, using eq. (9) [cf. footnote 1 below] we compute:

$$\begin{aligned}
E(\Sigma^2) &= \frac{1}{n-1}\sum_{i=1}^{n} \mathrm{Var}\left[\left(\frac{n-1}{n}\right) x_i - \frac{1}{n}\sum_{j\neq i} x_j\right] \\
&= \frac{1}{n-1}\sum_{i=1}^{n} \left[\left(\frac{n-1}{n}\right)^2 \mathrm{Var}(x_i) + \frac{1}{n^2}\sum_{j\neq i} \mathrm{Var}(x_j)\right] \\
&= \frac{1}{n-1}\sum_{i=1}^{n} \left[\left(\frac{n-1}{n}\right)^2 \sigma^2 + \frac{1}{n^2}\sum_{j\neq i} \sigma^2\right] \\
&= \frac{1}{n-1}\sum_{i=1}^{n} \left[\left(\frac{n-1}{n}\right)^2 \sigma^2 + \left(\frac{n-1}{n^2}\right) \sigma^2\right] \\
&= \frac{n}{n-1} \left[\left(\frac{n-1}{n}\right)^2 \sigma^2 + \left(\frac{n-1}{n^2}\right) \sigma^2\right] = \left(\frac{n-1}{n} + \frac{1}{n}\right)\sigma^2 = \sigma^2\,,
\end{aligned}$$

which completes the proof. Note that this computation justifies the presence of the denominator factor $n-1$ rather than n in eq. (4).

---

[1] Note that eq. (9) implies that $\mathrm{Var}(-x) = \mathrm{Var}(x)$. Hence, if x and y are independent random variables then it follows that $\mathrm{Var}(x-y) = \mathrm{Var}(x) + \mathrm{Var}(y)$.