Imagine we have a discrete quantum system (like spin) prepared in a state $|\psi_i\rangle$ and we measure a quantity A with possible values a_i.

$$\hat{A}\,|a_i\rangle = a_i\,|a_i\rangle$$

Now imagine we do a set of N measurements of A where each measurements yields a value a_{trial}, where n is the trial number.

For example, for a spin component of a spin-1/2 system, there are two possible values of spin component ($+\hbar/2$ and $-\hbar/2$) and there could be many trials where each trials takes on one of these values.

[Draw a diagram of a Stern-Gerlach measurement]

Remember, that the expectation value can be written in three different ways:

$$\langle \hat{A} \rangle = \frac{1}{N} \sum_{\text{trials}=1}^{N} a_{\text{trial}} \leftarrow \text{Good for tables of measurement data}$$

$$= \sum_i \mathcal{P}(a_i)\, a_i \leftarrow \text{Good for conceptual reasoning and probability histograms}$$

$$= \langle \psi_i | \hat{A} | \psi_i \rangle \leftarrow \text{Good for calculations given a state}$$

Expectation value is one characteristic of a distribution of measurement results. Another characteristic is how wide the distribution is. In statistics, this is called the standard deviation. In quantum mechanics, we call this the uncertainty. In this context, uncertainty does not mean there is an inprecision in the measurement process. Here, I mean there is an instrinsic distribution of measurement results due to the probabilistic nature of quantum systems.

Conceptually, the uncertainty is the average difference between a measured value and the average of the measurements. But if I use the usual way of calculating the average, by definition of the average, the result would be zero for any distribution. So, instead, I'll take a special average - the "root mean squared" or "RMS" average. The name tells you what to do. First, square the difference between a measured value with the average. Then take the average of these squares. Then take the square root.

$$\Delta A = \sqrt{\frac{1}{N} \sum_{trials} \left(\langle \hat{A} \rangle - a_{\text{trial}} \right)^2}$$

To use this definition, I need a set of data with N trials. What I'd like is a way to calculate the instrinsic uncertainty given a quantum state. To do this, I'll do some rearrangement. I start by expanding the square in the sum:

$$\begin{aligned}\Delta A &= \sqrt{\frac{1}{N}\sum_{\text{trials}}\left(\langle\hat{A}\rangle - a_{\text{trial}}\right)^2} \\ &= \sqrt{\frac{1}{N}\sum_{\text{trials}}\left(\langle\hat{A}\rangle^2 - 2\langle\hat{A}\rangle a_{\text{trial}} + a_{\text{trial}}^2\right)} \\ &= \sqrt{\frac{1}{N}\sum_{\text{trials}}\langle\hat{A}\rangle^2 - \frac{1}{N}\sum_{\text{trials}}2\langle\hat{A}\rangle a_{\text{trial}} + \frac{1}{N}\sum_{\text{trials}}a_{\text{trial}}^2}\end{aligned}$$

Looking at each term in the square root, I can make some simplifications. For the first term:

$$\frac{1}{N}\sum_{\text{trials}}\langle\hat{A}\rangle^2 = \frac{1}{\cancel{N}}\cancel{N}\langle\hat{A}\rangle^2 = \langle\hat{A}\rangle^2$$

For the second term:

$$\begin{aligned}\frac{1}{N}\sum_{\text{trials}}2\langle\hat{A}\rangle a_{\text{trial}} &= 2\langle\hat{A}\rangle \underbrace{\left(\frac{1}{N}\sum_{\text{trials}}a_{\text{trial}}\right)}_{\langle\hat{A}\rangle} \\ &= 2\langle\hat{A}\rangle^2\end{aligned}$$

For the third term:

$$\frac{1}{N}\sum_{\text{trials}}a_{\text{trial}}^2 = \langle\hat{A}^2\rangle$$

Notice that the square is inside the expectation value - the averaging happening after you square. Putting it all together:

$$\Delta A = \sqrt{\langle\hat{A}\rangle^2 - 2\langle\hat{A}\rangle^2 + \langle\hat{A}^2\rangle}$$

$$\Delta A = \sqrt{\langle\hat{A}^2\rangle - \langle\hat{A}\rangle^2}$$

This form is nice because I can calculate it given a state and an observable - I don't need data set of measurement trials.

One way to remember the order of terms under the square root is to notice that we need a real number for the uncertainty. $\langle\hat{A}^2\rangle$ is guaranteed to be positive, while $\langle\hat{A}\rangle$ is often zero. If $\langle\hat{A}\rangle$ is first, you might get an imaginary standard deviation, which you don't want.

Also notice that the uncertainty has the same dimensions as the observable.