**Where do counting errors come from?**

- Introduction

In the following note, I'd like to motivate the idea of an error on the number of decays you'd observe in some macroscopic amount of time. First, think about what I've been telling you in class:

$$\sigma_N = \sqrt{N} \qquad (1)$$

Just considering how you might have been dealing with errors in the past. If you measure $A$ with some error $\delta A$ and $B$ with some error $\delta B$, then if $C = A + B$, $\delta C = \sqrt{\delta A^2 + \delta B^2}$. Or, suppose you measure 9 events and then you measure 10 events in 2 equal time intervals. In one longer interval then you measured 19 events. The expected error on your measurement of 19 events is $\sqrt{19}$ which is $\sqrt{(\sqrt{10})^2 + (\sqrt{9})^2}$. So the prescription given in (1) above seems at least plausible.

- Randomness in Particle Decays

Now, lets consider the random nature of particle decays and write down some simple rules.

*a*) In a small interval of time, $\Delta t$, there is at most a single particle decay.

*b*) The probability of finding a decay in this interval of time is proportional to $\Delta t$.

*c*) A decay in $\Delta t$ is independent of decays at other times.

So, the probability of finding 1 decay in the time $\Delta t$ is:

$$P_1(\Delta t) = M\Delta t \quad (M = constant) \qquad (2)$$

and the probability of finding 0 decays in the time $\Delta t$ is:

$$P_0(\Delta t) = 1 - M\Delta t \quad (M = constant) \qquad (3)$$

And the probability of 1 *or* 0 decays is 1, as you'd expect. $P_1(\Delta t) + P_0(\Delta t) = 1$.

Now we're going to use a few tricks to try and get an expression for the probability of several decays occuring. It will rely on the rules above. Consider the probability of getting $N$ decays in a longer time interval $t + \Delta t$. As a calculational device, lets split this apart into the probability that we get $N$ decays in time $t$ and no events in a time $\Delta t$ later plus the probability that we get N-1 events in $t$ and 1 event in a time $\Delta t$ later.

$$P_N(t+\Delta t) = P_N(t)P_0(\Delta t) + P_{N-1}(t)P_1(\Delta t) \qquad (4)$$

Now substitute using (2) and (3):

$$P_N(t+\Delta t) = P_N(t)(1 - M\Delta t) + P_{N-1}(t)(M\Delta t) \qquad (5)$$

Re-arrange and form something that turns into a differential at very small $\Delta t$.

$$\frac{P_N(t+\Delta t) - P_N(t)}{\Delta t} = -MP_N(t) + MP_{N-1} \qquad (6)$$

or

$$\frac{dP_N(t)}{dt} = -MP_N(t) + MP_{N-1} \qquad (7)$$

This has a solution:

$$P_N(t) = \frac{1}{N!}(Mt)^N e^{-Mt} \qquad (8)$$

where:

$$\sum_{N=0}^{\infty} \frac{1}{N!}(Mt)^N e^{-Mt} = 1 \qquad (9)$$

- Operations with the Poisson Distribution

This is usually called the Poisson distribution, and we can do some calculating with it now. Consider the general expression for the most likely (or) avarage value of a quantity given a distribution:

$$< x >= \frac{\sum_a^b x f(x)}{\sum_a^b f(x)} \qquad (10)$$

Which in integral form looks like:

$$< x >= \frac{\int_a^b x f(x) dx}{\int_a^b f(x) dx} \qquad (11)$$

Try it. If $f(x) = 1$, this means that the distribution is flat. If you are trying to find the most likely value of $x$ between 1 and 2, the bottom integral is $2 - 1 = 1$ and the top integral is $2^2/2 - 1^2/2 = 3/2$, so $< x >= 3/2$ right in the middle. Which is what you expect.

To do calculations with (9) we're going to employ a dirty trick. Look at this expression:

$$\sum_{N=0}^{\infty} \frac{a^N}{N!} (Mt)^N e^{-Mt} = \sum_{N=0}^{\infty} \frac{1}{N!} (aMt)^N e^{-aMt} (e^{(aMt - Mt)}) = e^{(aMt - Mt)} \qquad (12)$$

$$\frac{d}{da} \sum_{N=0}^{\infty} \frac{a^N}{N!} (Mt)^N e^{-Mt} = \sum_{N=0}^{\infty} \frac{N a^{N-1}}{N!} (Mt)^N e^{-Mt} \qquad (13)$$

in the limit that we take $a = 1$ (13) becomes $< N >$.

$$< N >= \frac{d}{da} e^{(aMt - Mt)} |_{a=1} = Mt \qquad (14)$$

In counting decays, M is the average rate of decay (say 123.5/sec). So, if you measured for 1 sec., on average, you'd measure 123.5 decays, with a spread of values dictated by eqn. (8). Now, lets compute the spread on $< N >$. You've probably seen something called variance before. (I.e. for data distibuted like a bell curve, about 68 percent of the data is supposed to be within $\pm\sqrt{Variance}$ of the theoretical mean.) It is a measure of a spread in values.

$$Variance = \sum_{i}^{\#\ of\ measurements} \frac{(x_i - < x >)^2}{\#\ of\ measurements} \qquad (15)$$

Consider the case where we've taken a whole bunch of measurements, or trials, $N_i$, under the same conditions so that

$$Variance = \frac{\sum_{i}^{\# \ of \ trials} (N_i - < N >)^2}{\# \ of \ trials} \qquad (16)$$

$$Variance = \sum_{i}^{\# \ of \ trials} \frac{N_i^2 - 2 < N > N_i + < N >^2}{\# \ of \ trials} \qquad (17)$$

What is the most likely value for the variance?

$$< N_i > = < N > = Mt \qquad (18)$$

$$< N_i^2 > = (\frac{d^2}{da^2} + \frac{d}{da}) e^{(aMt - Mt)}|^{a=1} = (Mt)^2 + Mt \qquad (19)$$

$$< N_i^2 > -2 < N >< N_i > + < N >^2 = (Mt)^2 + Mt - 2(Mt)(Mt) + (Mt)^2$$
$$= Mt \qquad (20)$$

$$Variance = (\# \ of \ trials)(Mt)/(\# \ of \ trials) = (Mt) \qquad (21)$$

This means for a given measurement, you expect the value to be somewhere around $Mt$ with a spread given by (8) and characterized by (20). In a lot of cases, physicists will assume that the single measurement *is* the average so that you can compute an error. I.e. suppose For just one trial you know $M$ and $t$ beforehand, so you *expect* a measurement of $< N > = (Mt)$ events with an error of $\sqrt{Variance} = \sqrt{Mt}$. This is where equation (1) comes from. Now, if you took a bunch of measurements to find $(Mt)$, you need to modify the variance a little. (you can't really *measure* the spread (variance) with one trial unless you know Mt (the average value))

$$Variance = \frac{\sum_{i}^{\# \ of \ trials} (N_i - < N >)^2}{\# \ of \ trials - 1} \qquad (22)$$

- Applying the Poisson Distribution

Now it starts to get interesting. Suppose you were going to average a bunch of measurements you made. The expected spread on each one of the measurements you made should be determined by eqn. (15), but you need to use the theoretical mean that you don't know! What's up with that?!?!? What physicists typically do is cheat a little in order to get something that actually produces a servicable answer in a finite amount of time. We assume the measurement we took is close enough to the mean to give us a reasonable estimate of the error. This actually isn't too bad in a counting experiment if the number of events we took is around 10 or so. (Some references prefer 5, some 6, some 16 etc.) This means that when we go to compute a variance for a data set, we can estimate the error on each measurement by using the measurement. Lets examine a case in point, a fit to data where the end result is an average.

In order to perform a fit to data, notice the following. If we care to minimize the variance in equation in eqn. (15), we end up with an expression for the average:

$$\frac{dVariance}{dx_i} = \sum_i^{\#\ of\ measurements} \frac{2(x_i - <x>)}{\#\ of\ measurements} = 0 \qquad (23)$$

Which has the solution:

$$\begin{aligned}\sum_i^{\#\ of\ measurements} \frac{(x_i)}{\#\ of\ measurements} &= \sum_i^{\#\ of\ measurements} \frac{<x>}{\#\ of\ measurements} \\ &= \frac{\#\ of\ measurements <x>}{\#\ of\ measurements} \\ &= <x> \qquad (24)\end{aligned}$$

Notice something else about the equation for variance. Suppose that we know before hand the average and the variance (and for now we'll make the variance the same for each point $x_i$), we can re-arrange eqn. (15) to give us this expression.

$$\#\ of\ measurements \sim \sum_i^{\#\ of\ measurements} \frac{(x_i - <x>)^2}{Variance} \qquad (25)$$

So, if we wanted to check that our values for the average and the variance are reasonable, we could use eqn. (25) and see if we get a number that is close to the number of measurements. If we don't know the average before hand, eqn. (25) gets modified:

$$\# \ of \ measurements - 1 \sim \sum_{i}^{\# \ of \ measurements} \frac{(x_i - <x>)^2}{Variance} \qquad (26)$$

You can see that this is reasonable through the following evercise. Suppose that instead of just one average, you computed an average for every measurement you made. You'd end up with $x_i =< x_i >$ and the expression $(x_i - < x_i >)^2 = 0$ for every point, regardless of the variance associate with any single measurement. In this case, eqn. (26) would equal 0. Every time you use the data to find some variable, you need to increment eqn. (26) down by one. Usually, you see this written in the following form when you are trying to find $< x >$:

$$\chi^2 = \sum_{i}^{\# \ of \ measurements} \frac{(x_i - <x>)^2}{Variance_i} \qquad (27)$$

and

$$\frac{\chi^2}{\# \ of \ measurements - 1} \sim 1 \qquad (28)$$

and if you are trying to fit to a complicated function:

$$\chi^2 = \sum_{i}^{\# \ of \ measurements} \frac{(y_i - y(x))^2}{Variance_i} \qquad (29)$$

where

$$\frac{\chi^2}{\# \ of \ degrees \ of \ freedom} \sim 1 \qquad (30)$$

The number of degrees of freedom is just the number of measurements you made less the number of variables you are trying to find. For a fit to a straight line, for instance $y(x) = bx + a$, you are trying to find $b$ and $a$ so the number of degrees of freedom is 2. This makes sense too. It takes 2 points to make a straight line, but you need at least one more point to tell you if your straight line hypothesis is a good one (you can form a $\chi^2/d.o.f.$).

- ⋆ A more specific case of averaging

In the specific case of averaging where each measurement, $x_i$, we make has its own associated error, $\sigma_i$, we are trying to find the value $\overline{x}$ that minimizes the expression:(let $\# = \# \; of \; measurements$ for the sums below too)

$$\chi^2 = \sum_i^{\#} \frac{(x_i - \overline{x})^2}{\sigma_i^2} \qquad (31)$$

So we proceed:

$$\frac{d\chi^2}{dx_i} = 0 = \sum_i^{\#} \frac{2(x_i - \overline{x})}{\sigma_i^2} \qquad (32)$$

$$\sum_i^{\#} \frac{x_i}{\sigma_i^2} = \sum_i^{\#} \frac{\overline{x}}{\sigma_i^2} \qquad (33)$$

or

$$\frac{\sum_i^{\#} \frac{x_i}{\sigma_i^2}}{\sum_j^{\#} \frac{1}{\sigma_j^2}} = \overline{x} \qquad (34)$$

In order to find the error on our estimate of $\overline{x}$, we find how $\overline{x}$ varies with $\sigma_i$. In general, for a function $V = f(x, y, z)$, you find the error on $V$ by differentiating:

$$dV = \frac{\partial f(x,y,z)}{\partial x} dx + \frac{\partial f(x,y,z)}{\partial y} dy + \frac{\partial f(x,y,z)}{\partial z} dz \qquad (35)$$

Then you let $dV \sim \sigma_V$ and interpret the result like you are trying to find the length of a vector:

$$\sigma_V = \sqrt{\left(\frac{\partial f(x,y,z)}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f(x,y,z)}{\partial y}\right)^2 \sigma_y^2 + \left(\frac{\partial f(x,y,z)}{\partial z}\right)^2 \sigma_z^2 +} \qquad (36)$$

This can be more complicated when errors between variables become correlated. We'll skip this for now. So what do we have if we see how $\overline{x}$ varies with $x_i$?

$$\sigma_{\overline{x}} = \sqrt{\sum_{i}^{\#}\left[\left(\frac{\frac{1}{\sigma_i^2}}{\sum\limits_{j}^{\#}\frac{1}{\sigma_j^2}}\right)^2 \sigma_i^2\right]} = \sqrt{\frac{1}{\sum\limits_{j}^{\#}\frac{1}{\sigma_j^2}}} \tag{37}$$

Now I think you can tackle the note on exponential fitting on the web page.