STUDENT RESEARCH

Physics 133

An deepening of skills in the conduct of physics experimentation and, importantly, the documentation of experimentation.

Three experiments:

1) Atomic spectroscopy

2) AC circuits

3) Radioactive decay and shielding

[RADIATION SAFETY PRESENTATION
(by EH&S)]

NO FOOD / WATER

(20)

Our format: ~~18~~ classes

Class 1: Radiation safety
Overview and expectations
Beginning of statistics discussion } HW 1

Class 2: Completion of statistics lecture
Oscilloscope practice

Class 3: In class computing project } HW 2

Classes 4-8 LAB I [Report due class period ~~10~~ 9]

Classes ~~10-14~~ (10-13) 9-13 LAB II [Report due class period ~~15~~ 14]

Classes ~~15-19~~ (16-20), 14-18 LAB III [Report due ~~12/10 T Th~~ ~~12/9 MW~~ 3/16 at end of day.

You will be assigned to a lab group of 2.
There will be one group of 3 if necessary.
You may choose your lab partner but email me at baschumm@ucsc.edu by end of day tomorrow or you will be assigned one arbitrarily.

Course Web site:

~~http~~ scipp.ucsc.edu/~schumm and look for Ph133 link.

Expectations:

The writeup format is discussed on pp 7-13 of the lab manual. Some notes:

- You will work with a lab partner on acquiring data and thinking through the experiment, and perhaps analyzing the data (up to you).

- You will work on your own to generate the lab report

- The reports will be generated w/ a word processor (WORD, TeX, OpenOffice (caveat emptor) etc...), but HARD COPY will be handed in.

- Report Level: you are trying to tell someone w/ a similar background in physics to yourself what you did, how you did it, what you saw, and what it all means.

- Apparatus + Procedure really two separate sections. The latter should reference the former?

- You WILL NOT turn in your lab notebook, despite what the manual says. However, its advice on keeping the notebook is good. You will need to refer back to this when you write your report.

•) "Discussion + Conclusions" often the hardest part, even at a more advanced level
[Mr. Sonstorek story w/ litmus paper in 7th grade]

•) Plots: very important to learn how to generate effective titles, captions, + axis labels

•) You will be allowed to re-submit only the first report (but not to revisit the experiment itself)
It is expected that you will spend a significant amount of time outside of class on the reports

•) There will be programming (can be done w/ EXCEL (I will help), or whatever you like instead.

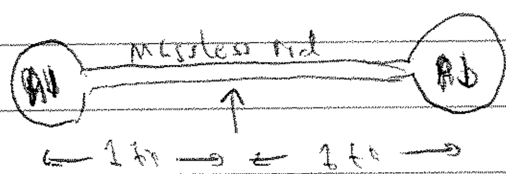
GRADING of 133

Hwork I	12.5%
Hwork II	12.5%
Lab I report	25%
Lab II report	25%
Lab III report	25%
	100%

Statistical Tools

"There are lies, there are damn lies, and there are statistics" - Benjamin Disraeli (?)

So let's learn some stats.
An important background notion is ~~It all begins with~~ weighted averages. → { mechanics, QM (exp values), Stat mech (ensembles), etc. }



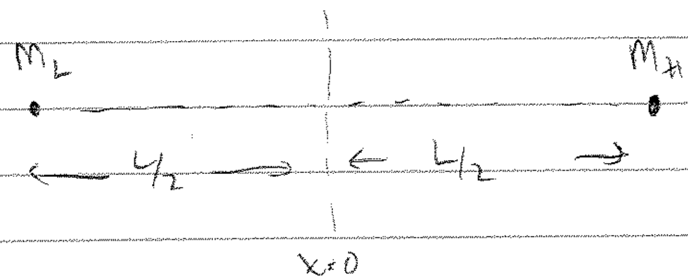
If you support this object right at its center, it will not balance.

How do we find the balance point? Or, in other words, where is "half-way" across the rod as far as the weights are concerned

⇒ what is the "weighted-average" position?

Consider a heavy + light (point-like) object separated by a distance L. Let the origin, for the sake of definiteness, be half-way between the two objects

*



Conjecture: the balance point will be at

$$X_B \stackrel{?}{=} \frac{(m_H)(x_H)+(m_L)(x_L)}{m_H+m_L} = \frac{(m_H)(L/2)+(m_L)(-L/2)}{m_H+m_L}$$

$$\cancel{= \frac{L}{2}\,\frac{m_H - m_L}{m_H +}}$$ Does this make any sense? Rewrite

$$X_B = \frac{L}{2}\,\frac{m_H - m_L}{m_H + m_L}$$ and explore some "obvious", or limiting cases

① $m_H = m_L \Rightarrow X_B = \frac{L}{2}\,\frac{0}{m_H+m_L} = 0$ ✓

② $m_L = 0 \Rightarrow X_B = \frac{L}{2}\,\frac{m_H}{m_H} = \frac{L}{2}$ ✓

③ $m_L = \frac{1}{2}m_H \Rightarrow X_B = \frac{L}{2}\,\frac{m_H - \frac{1}{2}m_H}{m_H + \frac{1}{2}m_H} = \frac{L}{2}\,\frac{\frac{1}{2}m_H}{\frac{3}{2}m_H} = \frac{L}{6}$,

or $\frac{1}{3}$ way from m_H to m_L. ✓

So that looks correct! What if we have a whole bunch of little masses spread out along a line?

weighted average position

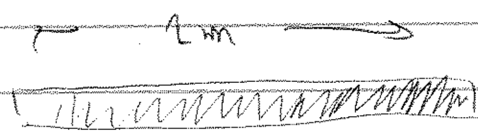
$$\langle x \rangle = \frac{m_1x_1 + m_2x_2 + \dots + m_nx_n}{m_1 + m_2 + \dots + m_n} = \frac{\sum_{i=1}^{n} m_i x_i}{\sum_{i=1}^{n} m_i}$$

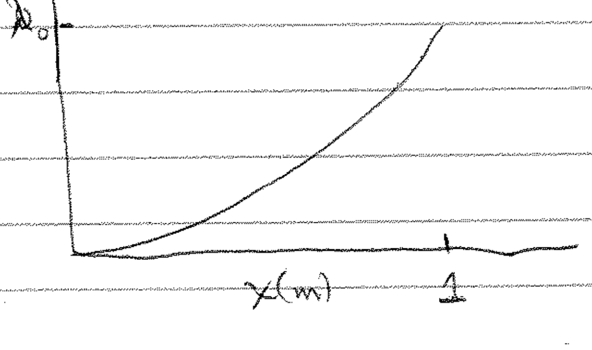
$$= \frac{\sum_{i=1}^{n} m_i x_i}{M} \quad \text{where } M \text{ is the total mass}$$

Next step: continuous distributions...

<u>The variable density rod</u>

Consider a rod that is very light on one end and dense on the other



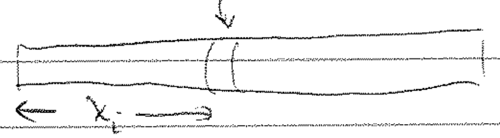


$\lambda(x) = \lambda_0 x^2$

$\boxed{S3}$

Where is the balance point (center of gravity) for this object? Consider a little slice dx of the rod located at x_i

$m_i = \lambda(x_i)\,dx$



$$\langle x\rangle = \frac{\sum x_i m_i}{\sum m_i} = \frac{\sum x_i \lambda(x_i)\,dx}{\sum \lambda(x_i)\,dx}$$

for $i = 1, \ldots, n$, where n is whatever it takes to cover the whole rod with dx's. But in the limit that $dx \to 0$, the sums just become integrals:

$$\langle x\rangle \to \frac{\int x\lambda(x)\,dx}{\int \lambda(x)\,dx} = \frac{\int x\lambda(x)\,dx}{M}$$

where M is the total mass of the rod. So, we just have a density-weighted average! Specifically, for

$$\lambda(x) = \lambda_0 x^2$$

$$\langle x\rangle = \frac{\int_0^1 x(\cancel{\lambda_0} x^2)\,dx}{\int_0^1 (\cancel{\lambda_0} x^2)\,dx} = \frac{\int_0^1 x^3\,dx}{\int_0^1 x^2\,dx} = \frac{\left[\frac{1}{4}x^4\right]_0^1}{\left[\frac{1}{3}x^3\right]_0^1} = \frac{3}{4}$$

Balance point is 3/4 way from 0-density end!

Poisson → Gaussian + Central limit theorem
→ Error prop. of Gaussian errors
→ How weighted avg. mean + error
χ^2

including points w/ different errors

Statistics

- Numerically, what have I observed?
- How accurate is that observation?
- If my observation is a function rather than a single number, how warranted is that functional form?

Easiest Case: ~~...~~ n equivalent measurements of x

$$\langle x \rangle = \frac{\sum_{i=1}^{n} x_i}{n} \quad \text{or (later)} \quad \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} 1}$$

is obvious guess of true value of x (e.g., my height measured 17 times)

How accurate a measurement? Actually two different questions you can ask

NATE SILVER!

1) If I do an 18^{th} measurement, how close is it likely to be to the true mean? STANDARD DEVIATION

2) How close is the combined 17-trial average to the true mean? ERROR ON THE MEAN

These are very different things and some nominally smart people have been confused by them.

[≈ like "ranking vs tasks" ?

[precision v. accuracy]

Standard Deviation

If you have 17 measurements, you can get an idea of how ~~accurate~~ precise each is just by comparing each w/ their mean. Maybe take an average deviation:

$$\text{Precision} = \text{"}s\text{"} \stackrel{?}{=} \frac{\sum (x_i - \langle x \rangle)}{n-1}$$

NOTE 1: We use $n-1$ since, e.g., if there's only 1 msmt then $x_1 = \langle x \rangle$ and $\sigma = 0$ which is clearly an underestimate. "You get one for free" with this "data-driven" approach to estimating the precision, so we need to compensate + divide by a smaller number.

NOTE 2: ~~Precision~~ By definition there are as many x_i's to left as to right of $\langle x \rangle$, so on average $x_i - \langle x \rangle \approx 0$. So no matter how ~~precise or imprecise~~ good or poor a measurement, the definition always gives $\sigma \sim 0$. No good!

So what we want, if we think about it, is a mean of a distance that is always positive, such as

$$s \stackrel{?}{=} \frac{\sum |x_i - \langle x \rangle|}{n-1}$$ mean absolute deviation

$$s = \sqrt{\frac{\sum (x_i - \langle x \rangle)^2}{n-1}}$$ square root of mean square deviation

[SD]

Both of these give some sort of mean deviation, or ~~mean~~ precision, in meters, or whatever unit you are trafficking in. Which is correct?

Answer: Both/neither. All depends. But for reasons we will see shortly, the latter is usually more applicable. But not always... but we'll stick solely w/ the latter.

Convenience for Calculation

If n is large, $\sqrt{n} \simeq \sqrt{n+1}$ and so [if not the correction to this is easy to make]

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \langle x\rangle)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(x_i^2 - 2x_i\langle x\rangle + \langle x\rangle^2\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}x_i^2 - 2\langle x\rangle\frac{1}{n}\sum_{i=1}^{n}x_i + \langle x\rangle^2\frac{1}{n}\sum_{i=1}^{n}1$$

$$= \langle x^2\rangle - 2\langle x\rangle\langle x\rangle + \langle x\rangle^2 = \langle x^2\rangle - 2\langle x\rangle^2 + \langle x\rangle^2$$

$$\boxed{s^2 = \langle x^2\rangle - \langle x\rangle^2}$$

Might want to remember this: most generalizable form of "variance" (QM uncertainty principle, etc.)

Practically, note that we can calculate means + variances by simply accumulating 3 sums as the data comes in:

$$\Sigma 1 \qquad \Sigma x_i \qquad \Sigma x_i^2$$
$$\overset{\shortparallel}{n}$$

and then

$$\langle x \rangle = \frac{\Sigma x_i}{\Sigma 1} \qquad s^2 = \frac{\Sigma x_i^2}{\Sigma 1} - \left(\frac{\Sigma x}{\Sigma 1}\right)^2$$

$$\langle x^2 \rangle = \frac{\Sigma x_i^2}{\Sigma 1}$$

Note that, surprisingly, you need not do two steps (with the first step being just calculating the mean)

Error on the Mean

So this standard deviation or variance tells how far off the mean a single measurement would be expected to be. But clearly, the more measurements you take and average together, the closer the average will be to the true mean!

Standard error $= s^2 = \langle x^2 \rangle - \langle x \rangle^2$ = how far off a single msmt is expected to be

But of course there is some sort of distribution of the measurements about the mean. More on distributions presently.

So, instead of 1 msmt, we take n msmts and average them: the averaging will lead to a more precise result than just a single msmt. How precise?

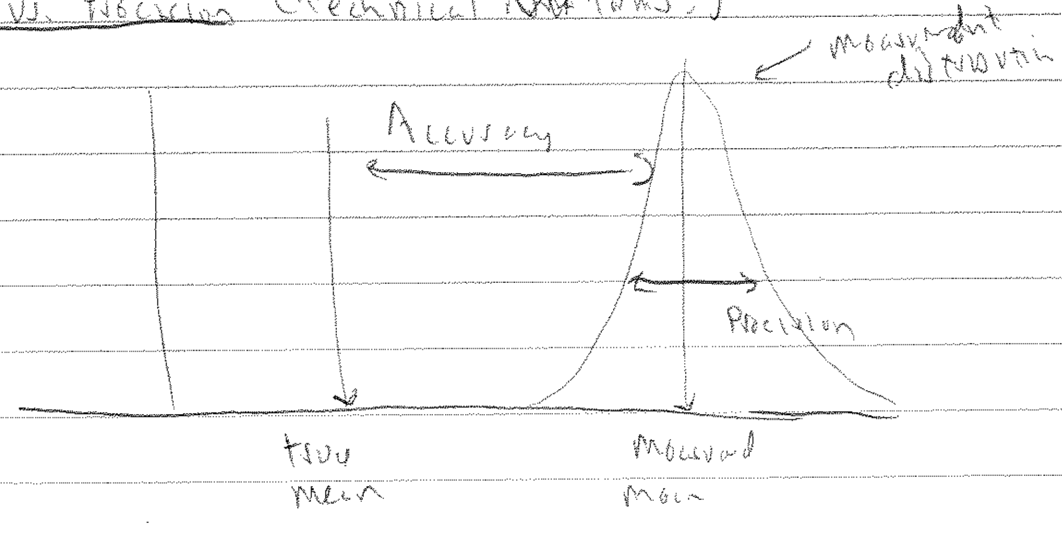
It can be shown that for a normal distribution of single msmts about the mean (more on this dist soon!)

$$S_{\bar{x}}^2 = \frac{S^2}{n}$$ ← error on single msmt, as above

error on the mean $\bar{x}$ of n measmts

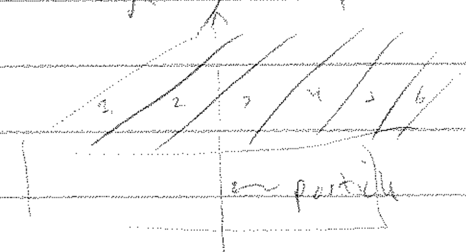
or $S_{\bar{x}} = \frac{s}{\sqrt{n}}$ ⇒ precision improves w/ sqrt of number of averaged trials

<u>Accuracy vs Precision</u> (technical ~~non~~ terms!)



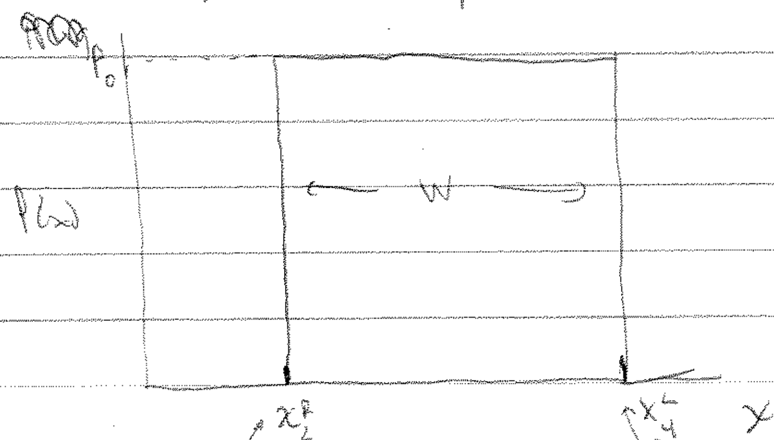
Probability Distributions

(a) Discrete systems: e.g. μ-strip detector



You hit one + only one of the strips. Let's say strip 3 is hit. You know the particle went somewhere between strips 2 and 4, but that's it.

$\Rightarrow$ Even, or "flat" probability density distribution



right edge of strip 2 — left edge of strip 4

So, prob. density distribution is

$$P(x) = \begin{cases} 0 & x < x_2^R \\ P_0 & x_2^R \le x \le x_4^L \\ 0 & x > x_4^L \end{cases}$$

and what is P_0? Well, particle had to go <u>somewhere</u>, so

$$1 = \int_0^\infty P(x)\,dx = \int_{x_2^R}^{x_4^L} P_0\,dx = P_0\left(x_4^L - x_2^R\right) = P_0 w$$

so $P_0 = 1/w$ where w = "strip width, or "pitch".

So, with the simplest of prob. density functions, there are some basic concepts.

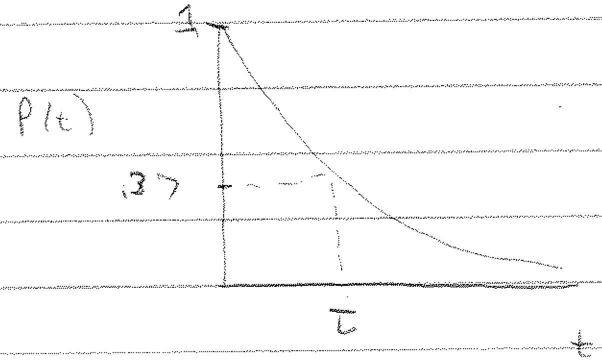
② Decaying systems

Consider a whole bunch of objects that decay, on average, with a mean lifetime of τ sec.

Look at 1 such object starting at some time that you call $t=0$.

Probability of finding the object un-decayed after a time t clearly falls as t grows. In fact, it can be shown to be

$P(t) = e^{-t/\tau}$ $\Rightarrow$ "exponential decay"



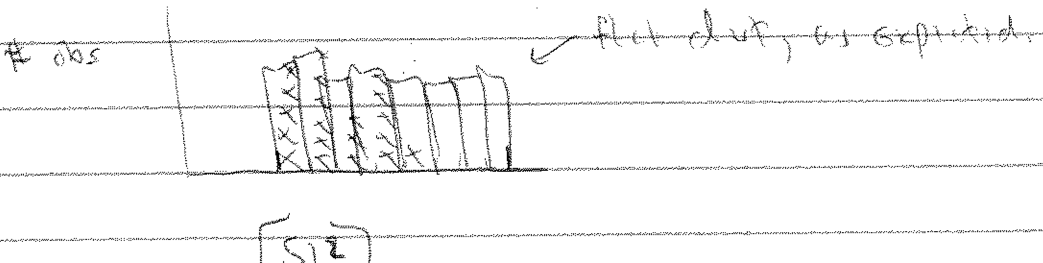
One of the two distributions most useful to us. The other is the

(3) Normal distribution [Confidence levels?]

The distribution that many processes tend to fall in.

"Central Limit Theorem" (mostly not a cocktail-party responsibility!)

→ Take any distribution, e.g., flat. Choose single samples. Histogram them (frequency distribution)



But now, each meas't is not a single sample, but a sum of two:

$$m_1 = x_1 + x_2$$
$$m_2 = x_3 + x_4$$
$$m_3 = x_5 + x_6$$

Distribution is no longer flat - do you see why? In fact, it will be triangular. For 3 ~~mea~~ summed samples, it starts to round out a bit.

In the limit of a large number of summed samples, per ~~approach~~ measurement, it approaches the normal or gaussian distribution

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

where it can be shown that

(A) $$\int_{-\infty}^{\infty} P(x)\,dx = 1 \quad \text{(normalized)}$$

(B) $$\langle x \rangle = \mu$$

and ~~the variance is σ^2~~

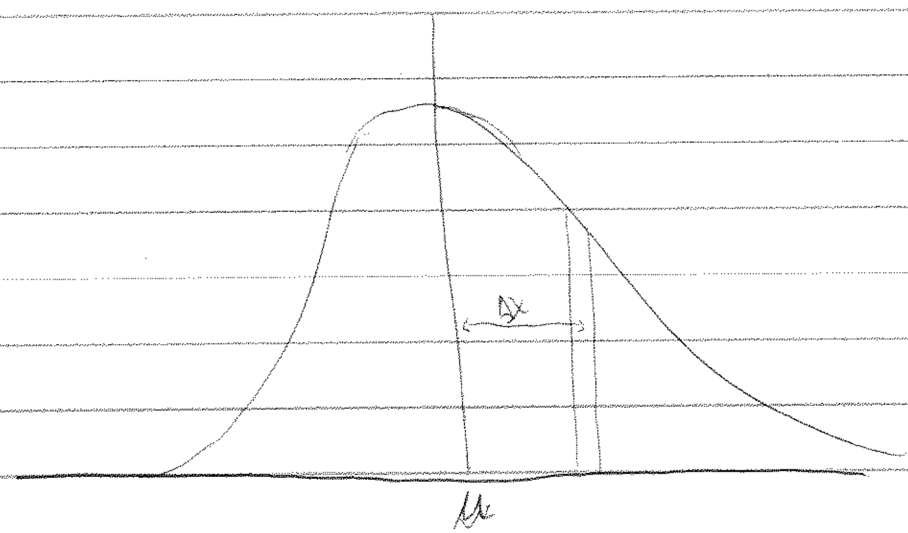
(C) $$\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2 = \sigma^2$$

This is the usual motivation for studying why we prefer the normal distribution (CL theorem); and these properties show why, given our preference for the normal dist, we choose to define the variance as we did several pages back.

OK - but don't trust me. How would we show this?

$1 = \int_{-\infty}^{\infty} p(x)\,dx$ $\Rightarrow$ cute "trick" to do this in closed form (ask me!?)

Now what about $\langle x \rangle$, σ^2. Let's say a word about this.



How precise is a measurement that is drawn from this distribution according to our

$$\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2$$

measure?

(5.14)

Consider contribution from slice between ~~μ+Δx and μ+Δx+dx~~ x and x+dx as shown above. Its contribution to the variance is prop to

$$\sim (\Delta x)^2 = (x-\mu)^2$$

but we also need to consider how much of the probability lies Δx from the mean μ. This is just $P(x)\,dx$, so the actual contribution is

$$(x-\mu)^2 P(x)\,dx$$

and when we sum over all such contributions and let $dx \to 0$ we get

$$s^2 = \frac{\int_{-\infty}^{\infty} (x-\mu)^2 P(x)\,dx}{\int_{-\infty}^{\infty} P(x)\,dx} \;\cancel{= \text{the same result}} = \sigma^2$$

Where the denominator is needed because the answer can't depend on the height of the dist. (and can be ignored if P(x) is normalized)

⇒ weighted (by prob density) mean squared deviation of prob density fnc.

Note also $\langle x \rangle \,\cancel{\mu} = \frac{\int x P(x)\,dx}{\int P(x)\,dx} \;\cancel{= \text{the same result}} = \mu$

Direct application again in QM!!

[S15]

Next: CL, then cumulative erf [illegible]

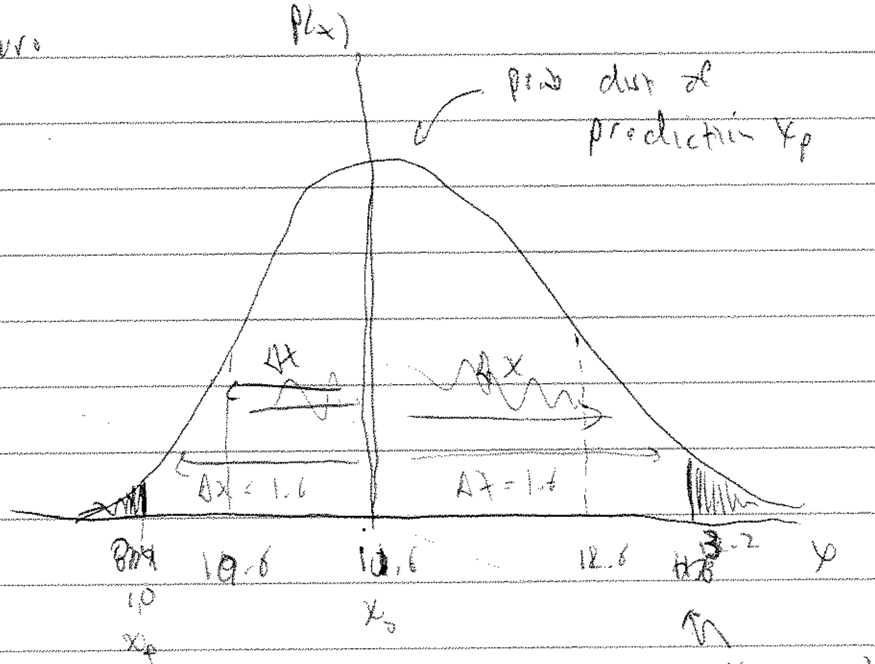
Gaussian Confidence Levels (CLs)

You predict $x_p = 10 \pm 1$ (e.g. # SM events in search for new physics). You measure $x_0 = 11.6 \pm 1$. How likely is it that your prediction is correct?

For current discussion, say that x_0 is $\bar{x}$, average over many, many msmts so that observed value is perfectly precise

Also, assume x_p uncertainty is gaussian distributed (assume from now on unless otherwise stated)

We have



Probability of getting exactly 11.6 again is exactly 0. How do you state the CL that the prediction works?

Usually: the symmetrized prob integral of the prob dist outside the measured difference:

$$CL = 1 - \int_{10.0}^{13.2} P(x)\,dx = 1 - \int_{x_0-\Delta x}^{x_0+\Delta x} P(x)\,dx$$

We can characterize the CL in terms of significance

$$\delta = \frac{\Delta x}{\sigma}$$

where Δx is the deviation. So for us, $\Delta x = 1.6$, $\sigma = 1$ and so

$$\delta = 1.6 = \text{"1.6 sigma deviation"}$$

Here are some important numbers

δ	CL
1	0.32
2	0.046
3	0.0027
4	0.000063

or

CL	δ
20%	1.28
10%	1.64
5%	1.96
1%	2.58
0.1%	3.29

$4\sigma \approx 3/10{,}000$

ERROR PROPAGATION

You may already be familiar with the "rules" associated with the propagation of errors through functions that relate your measurements to your results.

These are in fact not rules, but derivable properties under the assumptions that

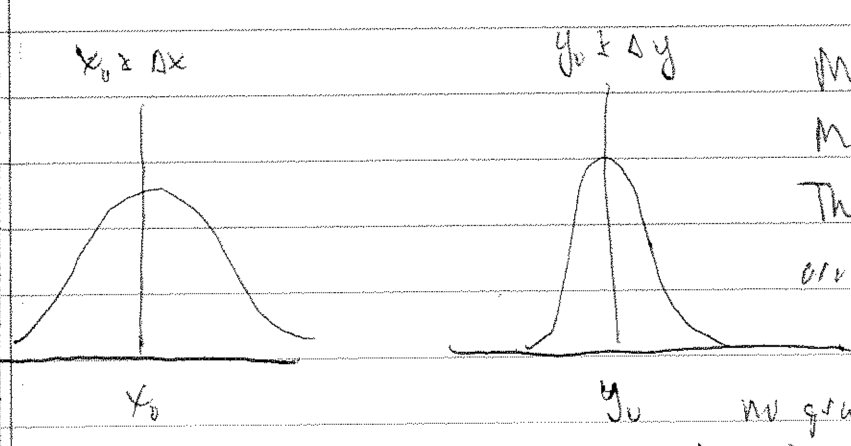
- ·) Uncertainties are gaussian
- ·) Uncertainties are relatively small, so that leading order treatment (Taylor expansion) can be used.
  1st term in

For the purposes of ~~this class~~ these lectures, we will also assume ~~that~~

- ·) All uncertainties are uncorrelated.

We'll talk a bit about correlated uncertainties in the context of ~~the atomic spectrum lab~~.

This need not be assumed, and generalizations are relatively straightforward once you understand the uncorrelated treatment. All we'll do for now is take a small digression & define what we mean by correlation...



Measure x once $\Rightarrow x_m$
Measure y once $\Rightarrow y_m$
The uncertainties Δx, Δy are uncorrelated if a measured $x_m > x_0$ leads to no greater (or lesser) prob. that $y_m > y_0$, and vice versa. (S.Y.)

Example: ~~salary~~ income, age

If person i has ~~high income~~ above avg. income they are also likely to have above average age ⇒ correlation!

Example: exo ~~planet~~ hint, age

Not correlated (to best of my knowledge!)

$\boxed{\text{End of Digression}}$

Error Combination

You are considering trying out for the olympics in the swim medley event ⇒ 50 m fly, 100 backstroke (separate events, plenty of time to rest). To make the trials, you need to beat a combined time of 1:50 = 110 s in a local time trial. You want to know if it's worth registering & paying the entree fee.

After many time trials, you have measured your times

50 m fly $45 \pm \overset{\sigma_f}{3}$ s $= \mu_f \pm \sigma_f$

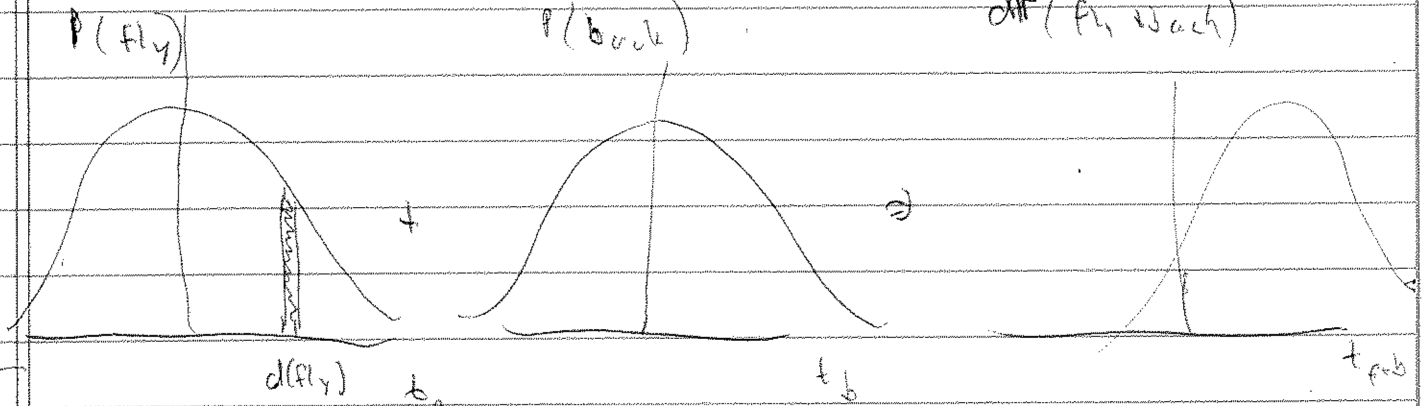
100 m back $70 \pm \overset{\sigma_b}{4}$ s $= \mu_b \pm \sigma_b$ What is "?", i.e. σ_{f+b}

Total $\underset{\mu_{f+b}}{115} \pm \underset{\sigma_{f+b}}{?}$ s

[19]

~~If these uncertainties are gaussian distributed,~~

There is a well-defined way to answer this:
convolution of probability distributions



For each d(fly), add complete "back" distribution to combined dist. D_i for each d(fly). (easy to write down expression, but not helpful)

IF the uncertainties are gaussian, then

Fly $$G_f(t_f) = \frac{1}{\sqrt{2\pi}\,\sigma_f}\, e^{-(t_f-\mu_f)^2/2\sigma_f^2}$$

back $$G_b(t_b) = \frac{1}{\sqrt{2\pi}\,\sigma_b}\, e^{-(t_b-\mu_b)^2/2\sigma_b^2}$$

and the above procedure can be shown to yield, mathematically,

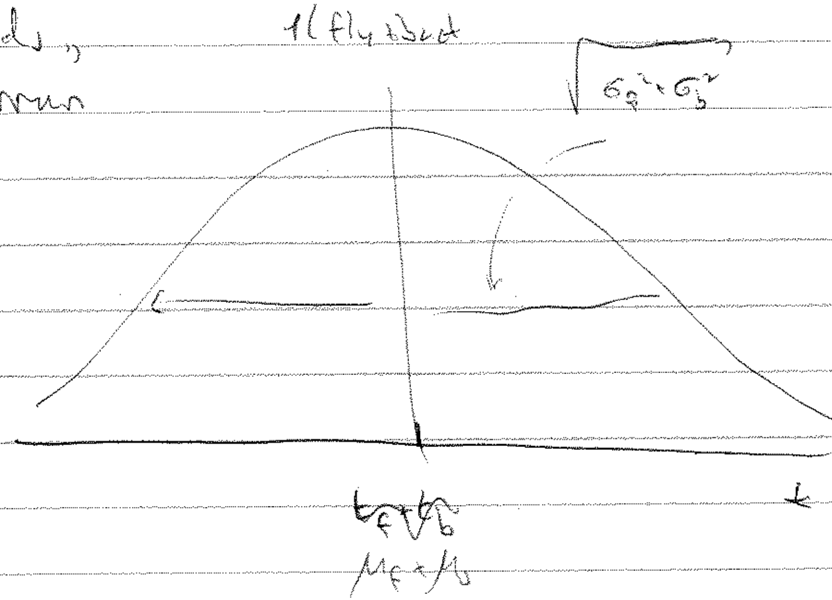
fly + back $$G_{f+b}(t) = \frac{1}{\sqrt{2\pi}\,\sqrt{\sigma_f^2+\sigma_b^2}}\, e^{-(t-[\mu_b+\mu_f])^2/2[\sigma_f^2+\sigma_b^2]}$$

for $t = t_f + t_b$

Or, in other words,
a distribution w/ a mean
of $t_f + t_b$ (good!)
but an uncertainty of

$$\sigma_{f+b} = \sqrt{\sigma_f^2 + \sigma_b^2}$$

"quadrature sum"!!



This is pretty much the underlying principle when it comes to error combination, driving rigorously for uncorrelated gaussian errors (and w/ a straightforward generalization for correlated errors).

So to reiterate:

A measurement of $a \pm \sigma_a$ added to a measurement of $b \pm \sigma_b$ yields

$$(a+b) \pm \left(\sqrt{\sigma_a^2+\sigma_b^2}\right)$$

So, in our example, time is 115 ± 5 seconds - within 1 SD of what you need!
How likely are you to qualify? CLs!! (chart)

[52]

Further Note: this is what's actually behind the $\sqrt{n}$ improvement in the error on the mean! To wit, take an average of two msmt ~~$X_1 + X_2$~~ $X_1 \pm \sigma_x$ $X_2 \pm \sigma_x$ of same dist:

$$\sigma_{x_1+x_2} = \sqrt{\sigma_x^2 + \sigma_x^2} = \sqrt{2\sigma_x^2} = \sqrt{2}\,\sigma_x$$

But average is $\frac{x_1+x_2}{2}$, so

$$\sigma_{x_1+x_2/2} = \frac{1}{2}\sqrt{2}\,\sigma_x = \frac{\sigma_x}{\sqrt{2}}$$

In fact for n msmts

$$\sigma_{x_1+\dots x_n} = \sqrt{n\sigma_x^2} = \sqrt{n}\,\sigma_x^2$$

$$\frac{\sigma_{x_1+\dots+x_n}}{n} = \frac{\sqrt{n}\,\sigma_x^2}{n} = \frac{\sigma_x^2}{\sqrt{n}}\ !$$

So OK...

[S21']

Error Propagation through Functions

Consider some function $f(x,y)$. How does f change when x and/or y change? (Note that we could have any number of ind. variables $f(x_1, \ldots, x_n)$ but starting w/ just two will introduce the main principles without having to get too abstract).

Recall from multivariate calculus the following (very sensible) relationship; ~~which~~

$$d f(x,y) = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy$$

change in f due to changes in x and y

~~change in~~ f rate of change of f wrt x, holding y constant

change in x

rate of change in f ~~holding~~ wrt y, holding x constant

change in y

This holds exactly for infinitesimal shifts dx, dy, so hence the assumption that the errors should be small.

So, if dx represents an uncertainty in x, then $dx \sim \sigma_x$. Similarly, $dy \sim \sigma_y$. How do we use this to get uncertainty σ_f?

Recall σ_x σ_y uncorrelated, so if ~~raising~~ dx, ~~and~~ dy, increases f, dy as likely to increase as decrease. How to think about this?

(S22)

Insight

~~Answer~~: Lets let $f(x,y) = x+y$. Then we already know what we should get

$$df(x,y) = 1\cdot dx + 1\cdot dy$$

and $\sigma_f = \sqrt{\sigma_x^2 + \sigma_y^2}$, or $\sigma_f^2 = \sigma_x^2 + \sigma_y^2$

So then, we see that df is just again a sum of $\frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy$ and so in the general case

$$\boxed{\sigma_f^2 = \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2}$$

Lets look at a very important special case

$$f(x,y) = xy$$

$$df(x,y) = \frac{\partial(xy)}{\partial x}dx + \frac{\partial(xy)}{\partial y}dy = y\,dx + x\,dy$$

$$\sigma_f^2 = y^2\sigma_x^2 + x^2\sigma_y^2$$

But we can then divide through by $f^2(x,y) = x^2y^2$ to get

$$\frac{\sigma_f^2}{f^2} = \frac{y^2 \sigma_x^2}{x^2 y^2} + \frac{x^2 \sigma_y^2}{x^2 y^2} = \left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2$$

$$\boxed{\frac{\sigma_f}{f} = \sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}}$$

So, important principles from special cases of addition, mult

⇒ When adding two msmts, combined absolute uncertainty is quadrature sum of individual absolute uncertainties

$$\sigma_{x+y} = \sqrt{\sigma_x^2 + \sigma_y^2}$$

⇒ When multiplying two msmts combined relative uncertainty is quadrature sum of individual relative uncertainties

$$\frac{\sigma_{xy}}{xy} = \sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$$

Remember these two special cases / principles. At the heart of a lot of error combo! Note that one might also consider subtraction and division. Let's look at subtraction and show that the result is the same as addition (you can argue division + mult are same)

$F(x,y) = x - y$

$df = \cancel{dx - dy}\ \frac{df}{dx}dx + \frac{df}{dy}dy = dx + (-dy)$

$$\sigma_f = \sqrt{\left(\frac{df}{dx}\right)^2 \sigma_x^2 + \left(\frac{df}{dy}\right)^2 \sigma_y^2} = \sqrt{\sigma_x^2 + \sigma_y^2}$$

So subtraction, division just like add, mult!

<u>Combining Separate Msmts of the Same Distribution</u>

The development in the lab manual says it all. This is however a critical principle so let's just reproduce this now.

Say that you perform measurements of the decay time τ of an isotope on two separate days. On day 1 you do n_1 identical msmts and on day 2 n_2. Each individual msmt has uncertainty σ_τ

$$\bar{\tau}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} \tau_j \qquad \sigma_1^2 = \frac{\sigma_\tau^2}{n_1}$$

$$\bar{\tau}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} \tau_j \qquad \sigma_2^2 = \frac{\sigma_\tau^2}{n_2}$$

Now, how to average together these? Since they're all similar, clearly we just want, letting $n = n_1 + n_2$

$$\bar{\tau} = \frac{1}{n}\sum_{k=1}^{n} \tau_k = \frac{1}{n_1+n_2}\left(\sum_{i}^{n_1} x_i + \sum_{j}^{n_2} x_j\right)$$

$$= \frac{1}{n_1+n_2}\left(n_1 \bar{\tau}_1 + n_2 \bar{\tau}_2\right)$$

But note that $n_1 = \frac{\sigma_\tau^2}{\sigma_1^2}$ and $n_2 = \frac{\sigma_\tau^2}{\sigma_2^2}$

Cancelling the σ_τ^2 everywhere (they appear in num & den of $\bar{\tau}$)

$$\boxed{\bar{\tau} = \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}\left(\frac{\bar{\tau}_1}{\sigma_1^2} + \frac{\bar{\tau}_2}{\sigma_2^2}\right)}$$

Let's look at this: again it's just a weighted average. The better something is measured (smaller σ) the more it should count in the average.

$\frac{1}{\sigma^2}$ is often called the weight w. In this nomenclature,

$$\bar{\tau} = \frac{1}{\underbrace{w_1 + w_2}_{w}}\left(w_1 \bar{\tau}_1 + w_2 \bar{\tau}_2\right) = \frac{\sum_i w_i \tau_i}{\sum_i w_i} \Rightarrow \text{just a weighted average!}$$

And finally, the total uncertainty is

$$\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n} = \frac{\sigma_x^2}{n_1+n_2} = \frac{\sigma_x^2}{\frac{\sigma_x^2}{\sigma_1^2}+\frac{\sigma_x^2}{\sigma_2^2}} = \frac{1}{\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}} = \frac{1}{w_1+w_2}$$

Again, very central principle in the whole business

NEXT: Least squares fitting + χ^2.

LEAST-SQUARES FITTING and the χ^2 TEST

~~Today~~ Q: I have a bunch of data and a model with some free parameters. What values of the parameter fit the model best? How certain are they? Do they support the model or not?

e.g. My data should make a straight line. What are the slope + intercept and how certain are they? How likely is it that the data actually make a straight line?

Let's start with the simplest model: a bunch of (t = true) different msmt $x_i \pm \sigma_i$ of some single quantity x_t ~~(t = true)~~. In fact let's make it even easier: just two msmt. What is the best fit to the assumption x_1, x_2 are measuring the same quantity? The ~~mean~~ (weighted) mean, right? With the uncertainty just the uncertainty on the mean. How do we tell if x_1, x_2 really are measuring the same thing?

Averaging Data Revisited

In general, when you make a measurement of a quantity (such as x_t), your msmt x_m is not exactly equal to x_t due to msmt error. We expect it to be off by an amount that 68% of the time will be within σ_x, but sometimes it will be more or less.

We ~~can~~ ^have^ defined the significance of a measurement x_i as being

$$s_i = \frac{(x_i - x_t)}{\sigma_i}$$

For two measurements we could define a total significance $s = s_1 + s_2$, but somehow if x_1 and x_2 are both way off of x_t in the opposite direction we would not want the total significance to be 0. So instead define a squared significance

$$\chi^2 = \sum_i s_i^2$$

also called a "chi-squared"

For our two-meas. problem,

$$\chi^2 = s_1^2 + s_2^2 = \frac{(x_1 - x_t)^2}{\sigma_1^2} + \frac{(x_2 - x_t)^2}{\sigma_2^2} = w_1(x_1 - x_t)^2 + w_2(x_2 - x_t)^2$$

where we have just reminded ourselves that we have previously define $w_i = \frac{1}{\sigma_i^2}$ for purposes of ease + clarity.

Now, in our problem, what we know is x_1 + x_2, and what we want to find is x_t.

Idea of Least Squares Fitting

The most likely value of the sought-after parameter is the one that minimizes χ^2

i.e., minimized the overall significance of the differences of the measurements from the estimated true value.

Let's run with this.

First, to make life easy, recall our convenience (transparency) definition

$$\frac{1}{\sigma_i^2} = w_i$$

so now we can write

$$\chi^2 = w_1 (x_1 - x_t)^2 + w_2 (x_2 - x_t)^2$$

and we minimize χ^2 wrt the parameter x_t in the standard fashion:

$$0 = \frac{d\chi^2}{dx_t} = 2w_1 (x_1 - x_t)^2 + 2w_2 (x_2 - x_t)^2 =$$

$$0 = W_1 X_1 + W_2 X_2 - (W_1 + W_2) X_t$$

And on solving for X_t

$$X_t = \frac{W_1 X_1 + W_2 X_2}{W_1 + W_2} = \frac{\sum_{i=1}^{2} W_i X_i}{\sum_{i=1}^{2} W_i} \Rightarrow \text{Weighted average}$$

If this went immediately further, as $W_i \rightarrow \frac{1}{\sigma_i^2}$

$$X_t = \frac{\frac{X_1}{\sigma_1^2} + \frac{X_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}$$

just as we argued before should be the average. So the reasoning: it looks as if our "method of least squares" is somehow an optimal algorithm for estimating parameters

But even more so note that if we are one S.D. away from the mean w/ any given mean ($X_1 = X_t + \sigma_x$, $X_2 = X_3 = \dots = X_t$)

$$S^2 = \frac{(X_1 - X_t)^2}{\sigma_x^2} = \frac{(X_t + \sigma_x - X_t)^2}{\sigma_x^2} = \frac{\sigma_x^2}{\sigma_x^2} = 1$$

so "one standard deviation of" means a significance of 1.

[53]

Before, we said that for an ~~combo~~ avg of two msmts w/ uncertainties σ_1 and σ_2

$$\frac{1}{\sigma_A^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \quad \text{or} \quad W_A = \cancel{w_1 + w_2}$$

$$\sigma_A^2 = \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} = \frac{1}{W_1 + W_2} \Rightarrow \sigma_A = \frac{1}{\sqrt{W_1 + W_2}}$$

(square total significance!!)

So, what is the change in χ^2 corresponding to a mismeasurement by $\Delta x = \sigma_A = \frac{1}{\sqrt{W_1 + W_2}}$, which before we had claimed was 1 S.D. based on playing w/ the properties of gaussians? There are any number of ways to measure x_1 and x_2 so that the (weighted) avg ~~xme~~ is $x_t + \sigma_A$, but one way is to have both x_1 and x_2 be off by this amount

$$x_1 = x_2 = x_t + \sigma_A \qquad \text{Let's try this}$$

In that case

$$\chi^2 = \frac{(x_1 - x_t)^2}{\sigma_1^2} + \frac{(x_2 - x_t)^2}{\sigma_2^2} = W_1 (x_t + \sigma_A - x_t)^2 + W_2 (x_t + \sigma_A - x_t)^2$$

$$= W_1 \sigma_A^2 + W_2 \sigma_A^2 = \frac{W_1}{W_1 + W_2} + \frac{W_2}{W_1 + W_2} = \frac{W_1 + W_2}{W_1 + W_2} = 1$$

and so a difference in total significance of 1 ~~can~~ again corresponds to a 1 SD (32% likely) difference! So it all fits together!

[S32]

Goodness of Fit & The Chi-Sq Test

Note that we pick the value of x_t that minimizes the χ^2... but what is that minimum?

It wholly depends on how close x_1 and x_2 are to one another!

For the sake of argument, say $\sigma_1 = \sigma_2 = \sigma_x$. Then, our estimate for x_t will clearly be

$$x_t = \frac{x_1 + x_2}{2}$$

and the minimum χ^2 will be

$$\chi^2 = \frac{(x_1 - x_t)^2}{\sigma_x^2} + \frac{(x_2 - x_t)^2}{\sigma_x^2} = \frac{\left(\frac{x_2 - x_1}{2}\right)^2}{\sigma_x^2} + \frac{\left(\frac{x_1 - x_2}{2}\right)^2}{\sigma_x^2}$$

$$= \frac{(x_1 - x_2)^2}{2\sigma_x^2}$$

But this is as it should be: if x_1 & x_2 are w/in a fraction of a σ_x of one another, they are consistent w/ being meas. of the same number, i.e., consistent w/ the hypothesis. If they are far apart, they are inconsistent.

So, the minimum χ^2 is a "goodness of fit" indicator.

In fact, we can turn it into a % probability... hang on...

[Pursue Oddly]

The General Least Squares Method

Other times we want to fit parameters of a function

i) Measure decay rate vs. t and fit exponential

ii) Measure y and X and fit line, parabola, etc.

Let (x_i, y_i) be a set of measurements. Assume x_i known (time on a stop watch) and y_i has gaussian errors σ_{y_i}

Assume we have hypothesis that $y = f(x; a, b, c, \ldots)$ where a, b, c are some parameters, e.g.

$y = mx + b$

Hypothesis is that y is linearly related to x, but and if so, we want to find slope + intercept.

~~avg~~ Least-squares-fit $\Rightarrow$ ~~minimize χ^2 to find a, b~~ find a, b that minimize χ^2

χ^2 test $\Rightarrow$ minimum value of χ^2 tells you whether hypothesis is correct.

So, from χ^2, ...

[534]

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - f(x_i; a, b, \ldots))^2}{\sigma_i^2}$$

and note that $\chi^2(a, b, \ldots)$ is a function of the parameters you want.

Now all you need to do is find the values of the parameters $a, b, \ldots$ that minimize χ^2

i) If you can,

$$0 = \frac{\partial \chi^2}{\partial a}$$

$$0 = \frac{\partial \chi^2}{\partial b}$$

$$\vdots$$

System of equations

ii) If you can't

Algorithm to search "χ^2 space" and find minimum ("MINUIT" is very popular program

Then...

•) Let $a_{min}, b_{min}, \ldots$ be values that minimize χ^2

$$\chi^2_{min} = \chi^2(a_{min}, b_{min}, \ldots)$$

Then, let a_1 be value of a such that

$$\chi^2(a_1, b_{min}, c_{min}) = \chi^2_{min} + 1$$

Then, $a_1 - a = \sigma_a$ = standard error on a.

•) The value of χ^2_{min} tells you the probability that the hypothesis $y = f(x)$ is correct. How?

Digression: Degrees of Freedom

~~We have N msmt (x_i, y_i) $i = 1, n$~~

Ore

Go back to case that hypothesis is single number, what if we have 3 msmts instead of 2?

$$\chi^2 = \frac{(x_1 - x_t)^2}{\sigma_1^2} + \frac{(x_2 - x_t)^2}{\sigma_2^2} + \frac{(x_3 - x_t)^2}{\sigma_3^2}$$

$\Rightarrow$ χ^2 grows w/ # msmts regardless of whether hypothesis is correct! So we must adjust it for the number of msmts.

But also, the ~~more~~ more "free parameters" we have, the easier it is to make the function fit the msmnt.

But there's a rigorous (& beautiful!) treatment of all this

Let n be # msmts

Let m be number of function parameters (2 for line, 3 for parabola, etc)

~~#DOF~~ ν = "# of degrees of freedom" = $n-m$

Then, our goodness of fit measure is

"Chisq per DOF" = χ^2_{min}/ν

From this, we can calculate a probability the hypothesis is correct.

Is just basic ~~sig~~ gaussian significance for $\nu=1$, but otherwise you need a sophisticated function!

Finally: The Poisson Dist

Final Topic: ~~Some~~ other important probability dists (PDFs)

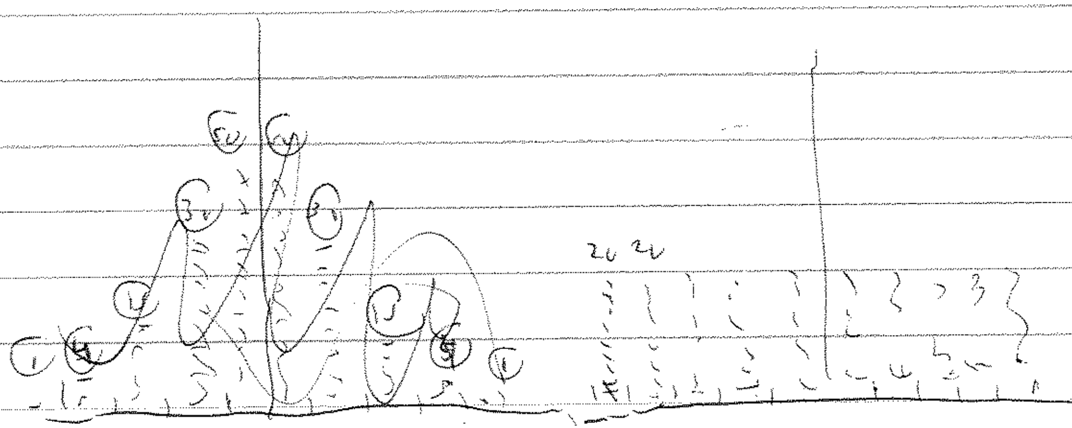
- Poisson
- Binomial

No theory... just the practical side

Poisson

Take any PDF $P(x)$ and do an experiment that samples from it. Record your results in a histogram: divide x axis into bins and record result in proper bin.

For example, sample a ~~gaussian~~ flat dist. What you won't see is



Why not? Because sampling is a statistical process, and you will fill up the bins randomly (random errors!)

Let's say after 200 samples, a perfect ~~gaussian~~ flat dist would give you the expected #s of counts shown in each bin.

In fact, you will not get ~~exactly~~ 20 in the second bin — something close, but different by the counting error.

What is it that describes the relative probability of getting say, 17 counts when you expect 20? Or, say, n counts when you expect m?

(538)

Answer: the Poisson distribution

Let's say that, on average, μ events will be seen in a given experiment.

Note that μ need not be integer... in our prior example, if the expt was 125 samples, $\mu = 12.5$ would be expected in bin 2..

Then, the probability of seeing n events in the experiment (bin 2) is

$$P(n,\mu) = \frac{\mu^n e^{-\mu}}{n!}$$

The transparency shows this distribution, but note that it is only meaningful for integer n.

~~Some properties of $P(n,\mu)$~~: The transparency shows $P(n,\mu)$ for $\mu = 1, 5, 10$ (they use k not n..)

Some properties of $P(n,\mu)$...

Normalization ~~Mean~~ of $P(n,\mu) = \sum_{n=0}^{\infty} P(n,\mu) = 1$

Mean of $P(n,\mu) = \sum_{n=0}^{\infty} n P(n,\mu) = \mu$

Variance^2 of $P(n,\mu) = \sum (n-\mu)^2 P(n,\mu) = \mu$

So, poisson dist has mean + variance of μ!

But note that it is asymmetric: $\mu^{+\Delta_+}_{-\Delta_-}$ w/ $\Delta_+ + \Delta_- = \mu$

is mean + ~~variance~~ range that contains 68% prob! ~~But how do you center~~

Finally, for large μ, $P(n,\mu) \to \frac{1}{\sqrt{2\pi}\,\mu} e^{-\frac{(n-\mu)^2}{2\mu^2}}$

ie, it becomes gaussian + symmetric (see transparency)

And finally, th binomial dist.

Say you are measuring an efficiency for someone to see a flash of light. You flash th light 100 times + record how many times she sees it.

$$\epsilon = \frac{\#\ \text{seen}}{100} \stackrel{?}{=} \frac{r}{100}$$

r ~ # successes

Lets say she's got good eyes + sees it 97 times

$$\epsilon_m = \frac{97}{100} = 0.970$$

But how certain? Well, counting stats says that uncertainty on 97 observations is $\sqrt{97}$, or 9.8

$$\Rightarrow \epsilon_m \stackrel{?}{=} 0.970 \pm 0.098$$

but this is nonsense! In fact, uncertainty is much closer to $\sqrt{3}$ - th counting stats on the number she didn't see. But if we take uncertainty in # failures instead, then what about low efficiencies?

We're think about it wrong.. in fact, # success prob. dist., given mean prob. of success ϵ and ~~m~~ trials, follows...

[54D]

Binomial Distribution

$$f(r; N, \epsilon) = \frac{N!}{r!(N-r)!} \epsilon^r (1-\epsilon)^{N-r}$$

where r = number ~~you see~~ of observed successes (the random variable, or "x", of the dist)

N = number of trials

ϵ = mean prob. of success

f = prob you will see r

w/ mean = $N\epsilon$ of course

variance $\sigma^2 = N\epsilon(1-\epsilon)$ (variance on r!)

Note: if $\epsilon = .97$, and you have 100 trials,

$\cancel{\sigma = \sqrt{100(.97)(.3)}}$ $\sigma_r = \sqrt{100(0.97)(0.03)} = \sqrt{2.91} = 1.71$

which is neither $\sqrt{97}$ nor $\sqrt{3}$ (but it's closer $\sqrt{3}$)

and $\sigma_\epsilon = \frac{\sigma_r}{N} = 0.0171$

It's also an asymmetric prob dist when ϵ is close to 1 or 0.