Probability

Discrete events:

Consider a random process which has a limited number of outcomes

tossing a coin: $n=2$

rolling a die: $n=6$

answering a SAT question: $n=4$

The outcomes have associated values $E_1, E_2, \dots E_n$

In each experiment (trial) only one outcome is observed.

Probability is defined as:

$$P(E_j) = \lim_{N\to\infty} \frac{N_j}{N}$$

N_j ← number of trials with outcome E_j

N ← number of trials performed

$$P_j = \lim_{N\to\infty} \frac{N_j}{N}$$

Examples: Coin: $E_1 = \text{heads}$

$E_2 = \text{tails}$

$$P(E_1) = P_1 = \frac{1}{2}$$

$$P(E_2) = P_2 = \frac{1}{2}$$

Die: $E_1 = 1 \quad E_2 = 2 \quad \dots \quad E_6 = 6$

$$P(E_1) = P(E_2) = \dots P(E_6) = \frac{1}{6}$$

Because $N_1 + N_2 + \dots N_n = N = \sum_{j=1}^{n} N_j$

(each trial had to result in one of the outcomes so total trials is sum of trials in each outcome)

That means we have limits on the value of each probability

$$0 \leq P(E_j) \leq 1$$

We also need to note that the combined probability of each event is 100%. (That is, there must be an outcome of each trial) This is called normalization:

$$\sum_{j=1}^{n} P(E_j) = \sum_{j=1}^{n} \lim_{N\to\infty} \frac{N_j}{N}$$

$$= \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{n} N_j$$

$$= \lim_{N\to\infty} \frac{N}{N} \qquad = 1$$

Now, consider an experiment in which two events A & B can or cannot occur. We define the outcomes as:

Possible Results	Number of occurrences
A and B	N_1
A, but not B	N_2
B, but not A	N_3
neither A nor B	N_4

We know right away that $N = N_1 + N_2 + N_3 + N_4$ and from the table we can deduce the probability of A or B separately)

$$P(A) = \frac{N_1 + N_2}{N} \qquad P(B) = \frac{N_3 + N_4}{N}$$

The Joint probability is the probability of A & B happening together

$$P(A, B) = \frac{N_1}{N}$$

Conditional probability, given that B occurs, what's the probability of A?

$$P(A|B) = \frac{N_1}{N_1 + N_3}$$

Similarly, probability of B given observation of A:

$$P(B|A) = \frac{N_1}{N_1 + N_2}$$

Playing around with these terms gives us a general result:

$$P(B|A)\,P(A) = \frac{N_1}{N_1+N_2} \cdot \frac{N_1+N_2}{N} = \frac{N_1}{N} = P(A,B)$$

$$P(A|B)\,P(B) = \frac{N_1}{N_1+N_3} \cdot \frac{N_1+N_3}{N} = \frac{N_1}{N} = P(A,B)$$

$$P(A \text{ given } B) \cdot P(B) = P(A \text{ and } B)$$

One interesting variant is what this implies for mutually exclusive A & B. That is if A precludes B, $P(A, B) = 0$

$$P(A|B) = P(B|A) = 0$$

For independent A & B A has no effect on appearance of B (or vice versa) so

$$P(B|A) = P(B)$$
$$P(A|B) = P(A)$$

$$\therefore \quad P(B|A)\,P(A) = P(B)P(A) = P(A,B)$$

product of independent probabilities

joint probability

Averages

Average roll of a die gives what?

$$\bar{X} = \langle x \rangle = \sum_j x_j \, P(x_j)$$

probability of outcome j

value of outcome j

$$= 1\cdot\tfrac{1}{6} + 2\cdot\tfrac{1}{6} + 3\cdot\tfrac{1}{6} \cdot 4\cdot\tfrac{1}{6} \cdot 5\cdot\tfrac{1}{6} + 6\cdot\tfrac{1}{6}$$

$$= \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5$$

note that this is not one of the outcomes!

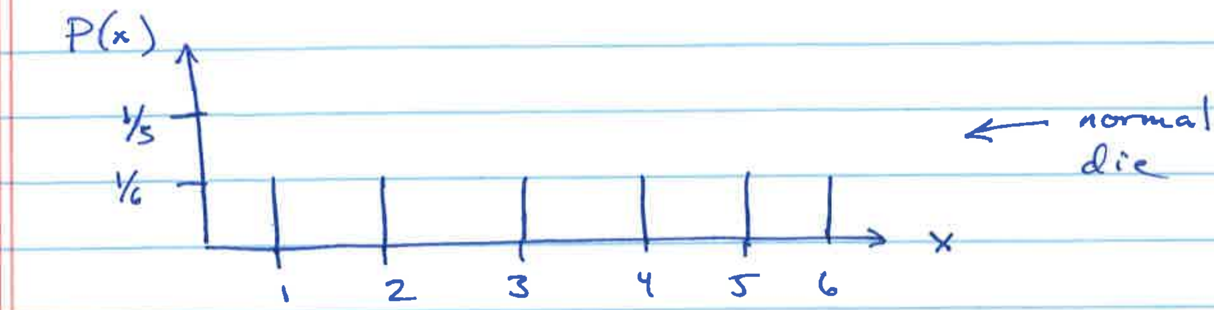
Loaded dice give very different means (we'll construct a die that gives 6 more often):

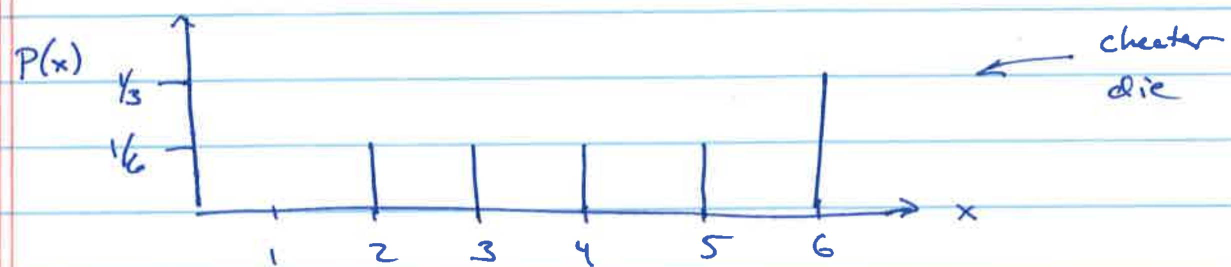
$$\langle x \rangle_{cheater} = 1\cdot(0) + 2\cdot\tfrac{1}{6} + 3\cdot\tfrac{1}{6} + 4\cdot\tfrac{1}{6} \cdot 5\cdot\tfrac{1}{6} + 6\cdot\tfrac{1}{3}$$

$$= \frac{2+3+4+5+12}{6} = \frac{26}{6} = 4\tfrac{1}{3}$$

We can plot $p(x_j)$ ← the probability of outcome x_j

vs. the value x_j





The second moment

$$\langle x^2 \rangle = \sum_j x_j^2 P(x_j)$$

average value of the <u>square</u> of the outcome

<u>Examples:</u> Normal die: $\langle x^2 \rangle = \dfrac{1+4+9+16+25+36}{6} = 15.1\overline{66}$

Cheater die: $\langle x^2 \rangle = \dfrac{4+9+16+25}{6} + \dfrac{36}{3} = \dfrac{27+36}{3}$

$= 21$

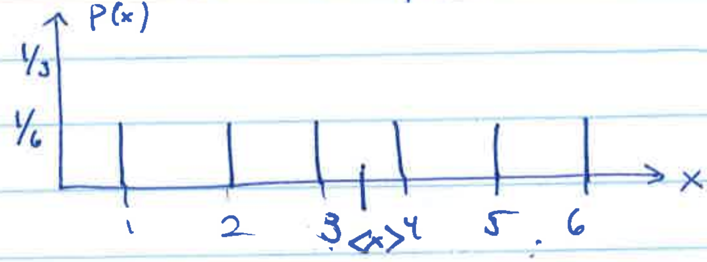
The standard deviation:

σ_x easiest to define as its <u>square</u>

$$\sigma_x^2 = \langle (x - \bar{x})^2 \rangle$$

Back to our dice example:

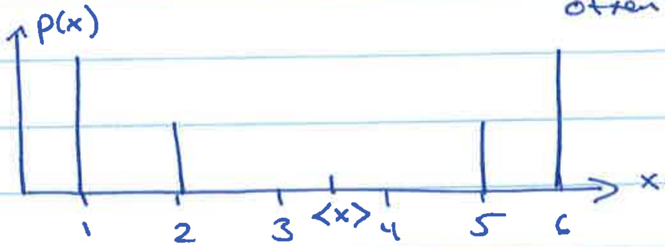
Normal :



Average

$$\langle x \rangle = \bar{x} = \sum_{j=1}^{n} x_j \, p(x_j) = \frac{1+2+3+4+5+6}{6} = 3.5$$

Cheater (super secret) : This die rolls 6 & 1 twice as often, but never rolls 3 or 4.



Average : $\langle x \rangle = 1\left(\frac{1}{3}\right) + 2\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{3}\right)$

$= 3.5$

Last time we defined 2 quantities

$$\langle x^2 \rangle = \text{second moment} = \sum_{j=1}^{n} x_j^2 \, p(x_j)$$

= the average value of the square of the dice roll.

For normal die: $\langle x^2 \rangle = \frac{1+4+9+16+25+36}{6} = 15.16\bar{6}$

For this cheater $\langle x^2 \rangle = \frac{1+36}{3} + \frac{4+25}{6} = 17.16\bar{6}$

Also the standard deviation σ_x

$$\sigma_x^2 = \langle (x - \langle x \rangle)^2 \rangle$$

Let's parse that

$$\sigma_x^2 = \langle (x - \langle x \rangle)^2 \rangle$$

The average square of the deviation from the average.

"On average how far are we (in absolute terms) from average behavior"

One of the things you'll prove in the HW is:

$$\sigma_x^2 = \langle (x - \langle x \rangle)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2$$

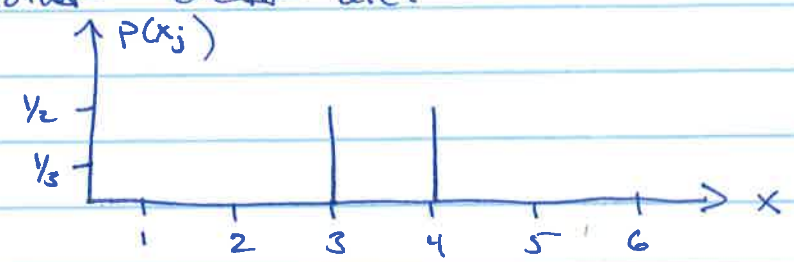
Here's a hint:

$$\sigma_x^2 = \sum_{j=1}^{n} (x_j - \langle x \rangle)^2 P(x_j)$$

So how does this work for the dice:

Normal	Super-secret cheater
$\langle x \rangle = 3.5$	$\langle x \rangle = 3.5$
$\langle x^2 \rangle = 15.16\overline{6}$	$\langle x^2 \rangle = 17.16\overline{6}$
$\sigma_x = \sqrt{15.16\overline{6} - (3.5)^2}$	$\sigma_x = \sqrt{17.16\overline{6} - (3.5)^2}$
$= 1.708$	$= 2.217$

Here's another cheater die:



Ok, you work this one:

$$\langle x \rangle = \tfrac{1}{2}\cdot 3 + \tfrac{1}{2}\cdot 4 = \tfrac{7}{2} = 3.5$$

$$\langle x^2 \rangle = \tfrac{1}{2}\, 9 + \tfrac{1}{2}\, 16 = \tfrac{25}{2} = 12.5$$

σ_x^2 in 2 ways:
$$\sigma_x^2 = \tfrac{1}{2}(3-3.5)^2 + \tfrac{1}{2}(4-3.5)^2$$
$$= \tfrac{1}{2}\left(\tfrac{-1}{2}\right)^2 + \tfrac{1}{2}\left(\tfrac{1}{2}\right)^2$$
$$= \tfrac{1}{2}\left(\tfrac{1}{4}\right) + \tfrac{1}{2}\left(\tfrac{1}{4}\right)$$

$$\sigma_x^2 = \tfrac{1}{4}$$
$$\sigma_x = 1/2$$

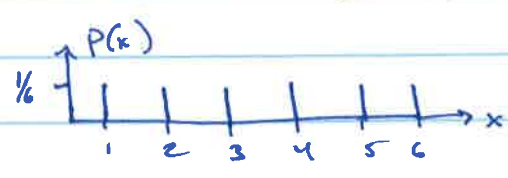
Also: $\sigma_x = \sqrt{12.5 - (3.5)^2} = \sqrt{\tfrac{1}{4}} = \tfrac{1}{2}$

So for this cheater, on average, we deviate from the mean (3.5) by a value of $\underline{1/2}$.

The more sharply peaked a distribution is around $\langle x \rangle$, the smaller σ_x will be!

Different kinds of distributions

A regular die has a uniform distribution



That is, for every possible outcome the probability is identical:

$$P(x_j) = \frac{1}{n} \leftarrow \text{total \# of outcomes possible.}$$

Another very common discrete probability distribution is called the Poisson distribution

$$P(j) = \frac{a^j}{j!} e^{-a}$$

where $j = 0, 1, 2, 3, \ldots$

and a is a constant.

This distribution governs discrete events that occur during a fixed time period (i.e. # of text messages you receive per day, # of mutations on a DNA strand / month, cars arriving at a traffic light, # of pieces of mail / day)

It is a shockingly common law at all length & time scales.

First, let's show it is normalized:

$$\sum_{j=0}^{\infty \leftarrow \text{all possible outcomes}} P(j) = \sum_{j=0}^{\infty} \frac{a^j}{j!} e^{-a}$$

$$= e^{-a} \sum_{j=0}^{\infty} \frac{a^j}{j!} = e^{-a}\left(1 + a + \frac{a^2}{2!} + \frac{a^3}{3!} + \ldots\right)$$

$$= e^{-a}\left(e^{a}\right) = 1$$

What about the average?

$$\langle j \rangle = \sum_{j=0}^{\infty} j \, P(j)$$

$$= \sum_{j=0}^{\infty} j \frac{a^j}{j!} e^{-a}$$

$$= e^{-a} \sum_{j=0}^{\infty} \frac{j \, a^j}{j!}$$

$$= e^{-a} \sum_{j=0}^{\infty} \frac{a^j}{(j-1)!}$$

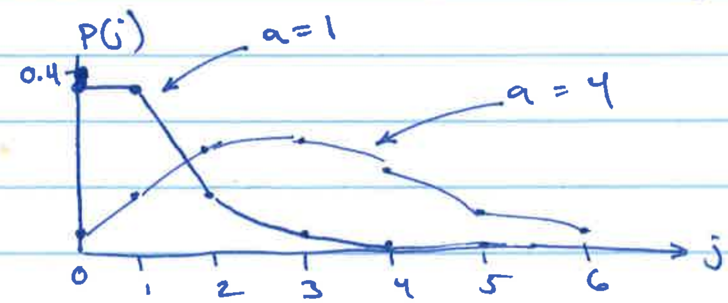
$$= e^{-a} \sum_{j=0}^{\infty} a \frac{a^{j-1}}{(j-1)!}$$

($(-1)! = \infty$) so the first term goes away!)

$$= a \, e^{-a} \, e^{a}$$

$$= a$$

This distribution is fascinating



We can generalize a few of the quantities we've defined:

n^{th} <u>moment</u> : $\langle x^n \rangle$

n^{th} <u>central</u> moment : $\langle (x - \langle x \rangle)^n \rangle$

1^{st} central moment: $\langle (x - \langle x \rangle) \rangle$

$$= \langle x \rangle - \langle\langle x \rangle\rangle$$
$$= \langle x \rangle - \langle x \rangle$$
$$= 0$$

2^{nd} central moment = variance : $\langle (x - \langle x \rangle)^2 \rangle$

$$\equiv \sigma_x^2 = \langle x^2 \rangle - \langle x \rangle^2$$

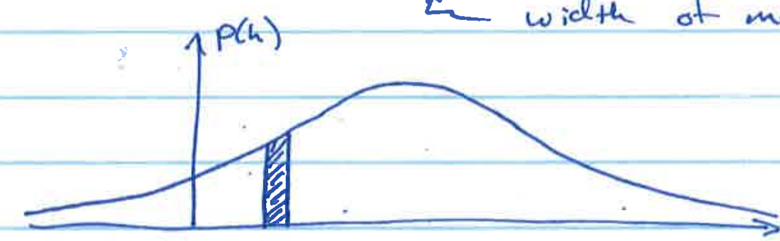
Since $\sigma_x^2 \geq 0$, it must be true that

$$\langle x^2 \rangle \geq \langle x \rangle^2$$

Continuous Distributions

Consider probabilities of finding a chemistry professor with a height between 5'3" & 5'4"

$p(h)\underbrace{dh}$ ← width of measurement = 1"



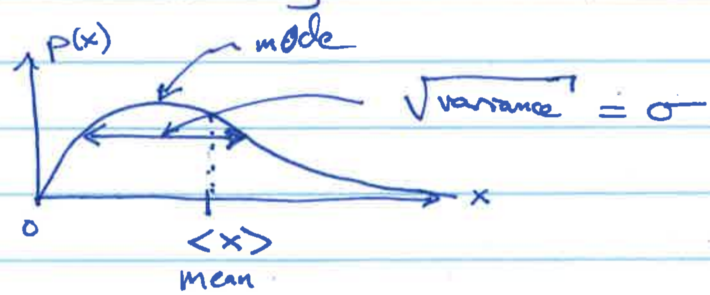
↑ $p(h)dh$ = probability any randomly-chosen ~~given~~ person has a height between 5'3" & 5'4"

Some analogies between dice (discrete outcomes) and height (continuous outcomes)

	Discrete	Continuous
Boundedness :	$0 \leq p_j \leq 1$	Not true: $0 \leq p(h) \leq 1$
Conservation :	$\sum_{j=1}^{n} p_j = 1$	$\int_{-\infty}^{\infty} p(h)dh = 1$
Mean	$\langle x \rangle = \sum_{j=1}^{n} x_j\, p(x_j)$	$\langle h \rangle = \int_{-\infty}^{\infty} h\, p(h)dh$

	Discrete	Continuous
2nd moment	$\langle x^2 \rangle = \sum_{j=1}^{n} x_j^2 P_j$	$\langle h^2 \rangle = \int_{-\infty}^{\infty} h^2 p(h)\,dh$
Std. Dev.	$\sigma_x^2 = \sum_{j=1}^{n} (x_j - \langle x \rangle)^2 P_j$	$\sigma_h^2 = \int_{-\infty}^{\infty} (h - \langle h \rangle)^2 p(h)\,dh$

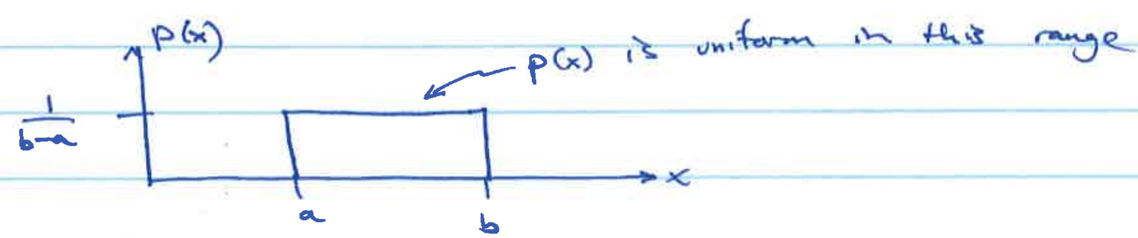
Consider an arbitrary distribution



Examples

Uniform probability density

$$p(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$



Normalization check:

$$\int_{-\infty}^{\infty} p(x)\,dx = \int_a^b \frac{1}{b-a}\,dx$$

$$= \frac{1}{b-a}\left[x\right]_a^b$$

$$= \frac{b-a}{b-a} = 1$$

Mean:

$$\langle x \rangle = \int_a^b x\, p(x)\, dx = \int_a^b \frac{x}{b-a}\, dx$$

$$= \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{b^2-a^2}{2(b-a)}$$

$$= \frac{(b+a)(b-a)}{2(b-a)} = \frac{b+a}{2}$$

2nd moment:

$$\langle x^2 \rangle = \int_a^b x^2\, p(x)\, dx = \int_a^b \frac{x^2}{b-a}\, dx$$

$$= \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b = \frac{b^3-a^3}{3(b-a)}$$

$$= \frac{b^2+ab+a^2}{3}$$

Variance:

$$\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2 = \frac{b^2+ab+a^2}{3} - \frac{(b+a)^2}{4}$$

$$= \frac{(b-a)^2}{12}$$

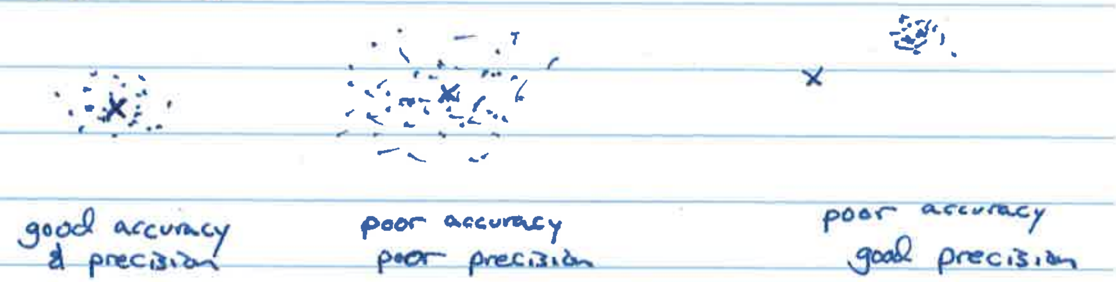
Next time: <u>error</u>

Treatment of Experimental Data

Types of error — All experiments contain sources of error, so it is important to determine the accuracy of measurements

Exact value: x

Measured values: •



- Systematic errors : same direction & magnitude for each repetition of the experiment: affect accuracy without affecting precision

- Random errors : do not have the same direction & magnitude each time: affects both accuracy & precision.

Statistical treatment of random errors

- Sample: A finite number of actual measurements
- Population : An imaginary set of an infinite number of repetitions of a perfect experiment.

We use samples to infer properties of populations.

Properties of a population

probability distribution of a property x : $f(x)$ N.B: $\int_{-\infty}^{\infty} f(x)\,dx = 1$

Population mean : $\mu = \int_{-\infty}^{\infty} x f(x)\,dx$

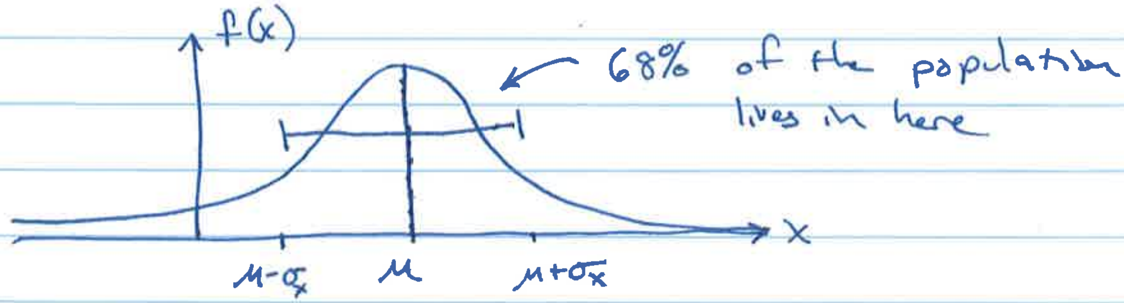
Population variance : $\sigma_x^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$

Population Std. Deviation : $\sigma_x = [\sigma_x^2]^{1/2}$

We can often make the assumption that any population of experimental results is governed by the Gaussian or Normal distribution:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}}\, e^{-(x-\mu)^2/(2\sigma_x^2)}$$

where μ = mean , σ_x^2 = variance



For 68.3% of the population $\mu - \sigma_x \le x \le \mu + \sigma_x$

For 95% of the population

$$\mu - 1.96\,\sigma_x \le x \le \mu + 1.96\,\sigma_x$$

The Central Limit Theorem: If a random experimental error arises from the sum of many contributions, then the errors will be governed by a Gaussian distribution.

Properties of a sample:

sample mean: $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$

i = individual experiments
N = total experiments
x_i = measurement of x in experiment i

sample variance: $S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$

sample std. dev: $S_x = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$

We use $N-1$ in the denominator to create an unbiased estimate of S_x.

N = N pieces of information = $N-1$ pieces of information in addition to the mean.

Scientists often report data such that the reported value will be correct 95% of the time. This requires the 95% confidence interval.

That is the correct value is

$$\mu = \bar{x} \pm \epsilon$$

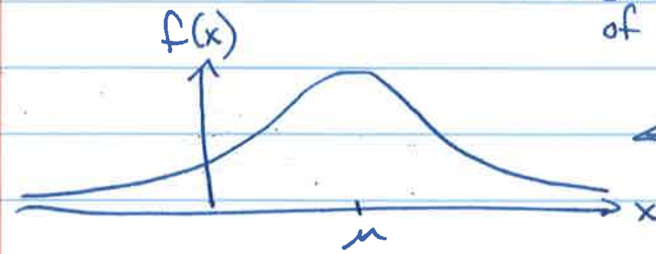
where 95% of the time μ lies between

$$\bar{x} - \epsilon \leq \mu \leq \bar{x} + \epsilon$$

If we knew σ, the population std. dev., we could say with 95% confidence that

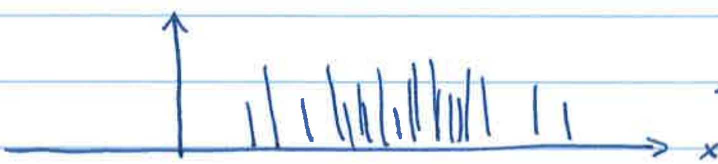
$$\mu = x \pm 1.96\,\sigma$$

x outcome of a single measurement of the population mean



original population with σ & μ

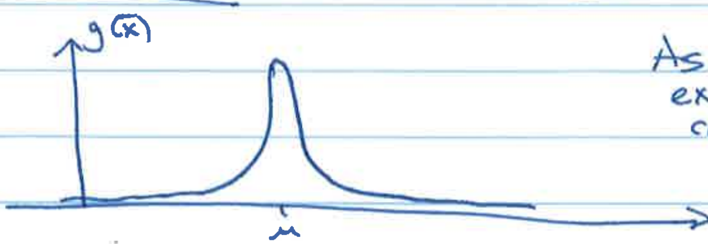
multiple experiments, each to measure mean value



values of $\bar{x}$ for N measurements

Standard Deviation of the measurements of the mean

$$\sigma_m = \frac{\sigma}{\sqrt{N}}$$



As we do more experiments we become confident that our mean-of-means $\bar{x}$ is closer to the correct value.

$$g(\bar{x}) = \frac{1}{\sqrt{2\pi\,\sigma_m^2}}\; e^{-(\bar{x}-\mu)^2/(2\sigma_m)^2}$$

Taking more measurements of a mean gives you less error in $\bar{x}$. Reducing error by 4 times requires 4 times as much data.

$$\epsilon = 1.96\,\sigma_m = \frac{1.96\,\sigma}{\sqrt{N}}$$

We still don't know σ, but

$$\epsilon = \frac{1.96\, S_x}{\sqrt{N}}$$

is a reasonable guess. The correct estimate of ϵ is

$$\epsilon = \frac{t(\nu, 0.05)\, S}{\sqrt{N'}}$$

where t is called "student's t-factor" and $\nu = N-1$.

Propagation of Errors

Consider a quantity that is calculated from 2 measured quantities

$$a = \bar{a} \pm \epsilon_a$$

← error confidence limits for a

↑ mean value of a

$$b = \bar{b} \pm \epsilon_b$$

$$c = \bar{a} + \bar{b}$$

How do we calculate the error in $\bar{c}$ from ϵ_a & ϵ_b?

More generally:

$$y = y(x_1, x_2, \ldots x_n)$$

$$x_i = \bar{x}_i \pm \epsilon_i \qquad (i = 1 \ldots n)$$

First, create differential:

$$dy = \left(\frac{\partial y}{\partial x_1}\right) dx_1 + \left(\frac{\partial y}{\partial x_2}\right) dx_2 + \dots \left(\frac{\partial y}{\partial x_n}\right) dx_n$$

finite size errors

$$\Delta y = \left(\frac{\partial y}{\partial x_1}\right) \Delta x_1 + \left(\frac{\partial y}{\partial x_2}\right) \Delta x_2 + \dots \left(\frac{\partial y}{\partial x_n}\right) \Delta x_n$$

Square this:

$$(\Delta y)^2 = \left(\left(\frac{\partial y}{\partial x_1}\right) \Delta x_1 + \dots \left(\frac{\partial y}{\partial x_n}\right) \Delta x_n\right)^2$$

The leading direct terms: there are others that couple the measurements too, but let's ignore for now.

$$(\Delta y)^2 \approx \left(\frac{\partial y}{\partial x_1}\right)^2 \Delta x_1^2 + \dots + \left(\frac{\partial y}{\partial x_n}\right) \Delta x_n^2$$

$$\epsilon_y \approx \sqrt{\left(\frac{\partial y}{\partial x_1}\right)^2 \epsilon_1^2 + \left(\frac{\partial y}{\partial x_2}\right)^2 \epsilon_2^2 + \dots}$$

As the errors $\epsilon_1 \dots \epsilon_n \to 0$ this becomes exact.

For $C = a + b$

$$\frac{\partial c}{\partial a} = 1 \qquad \frac{\partial c}{\partial b} = 1$$

$$\therefore \quad \epsilon_c = \sqrt{\epsilon_a^2 + \epsilon_b^2}$$

Example: Suppose we measure the frequency ω from an IR spectrum

$\omega \pm \epsilon_\omega \leftarrow$ 95% confidence limits

And we know a mass from mass spectrometry

$m \pm \epsilon_m$

Now, if we want the spring constant

$$k = m\omega^2$$

$$\frac{\partial k}{\partial m} = \omega^2 \qquad \frac{\partial k}{\partial \omega} = 2m\omega$$

$$\epsilon_k = \sqrt{\omega^2 \epsilon_m^2 + 2m\omega\, \epsilon_\omega^2}$$

95% confidence interval in spring constant.

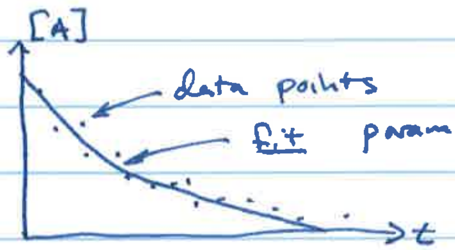
Linear Regression

Consider a function y of an independent variable x:

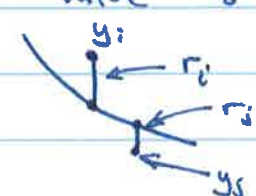
$$y = y(x) = f(x;\ \underbrace{a_1, a_2, \ldots a_p}_{\text{parameters or constants}})$$

We'd like the best set of parameters to fit a collection of data $\{x_i, y_i\}$ to a model

Example: $$[A] = [A_0]e^{-kt} = f(t;\ \underbrace{[A_0], k}_{\text{parameters}})$$



A residual for the i^{th} data point is the discrepancy between the measured values and the value of the function at that point:



prediction of fit

$$\underset{\text{residual}}{r_i} = \underset{\text{actual value}}{y_i} - \overbrace{f(\underset{\text{independent variable}}{x_i};\ a_1 \ldots a_p)}$$

A perfect fit would have $r_i = 0$ for all points i.
The best fit to the data (given a particular model) minimizes the sum of squares of the residuals

This is called the method of <u>least squares</u> or a regression.

That is:

$$R = \sum_{i=1}^{N} \left[y_i - f(x_i ; a_1 \ldots a_p) \right]^2$$

number of data points

We minimize R with respect to $\{a_1, \ldots a_p\}$

$$\frac{\partial R}{\partial a_1} = 0 \quad , \quad \frac{\partial R}{\partial a_2} = 0 \quad , \quad \ldots$$

This usally generates a set of nonlinear equations that must be solved iteratively, and the tools to do this (like Qt Grace) can require some expertise.

For <u>linear</u> functions like $y = mx + b$, the parameters m & b can be determined exactly.

$$R = \sum_{i=1}^{N} (y_i - mx_i - b)^2$$

$$\frac{\partial R}{\partial m} = 2 \sum_{i=1}^{N} (y_i - mx_i - b)(-x_i) = 0$$

$$\frac{\partial R}{\partial b} = 2 \sum_{i=1}^{N} (y_i - mx_i - b)(-1) = 0$$

Rewriting these:

$$m \sum_{i=1}^{N} x_i^2 + b \sum_{i=1}^{N} x_i = \sum_{i=1}^{N} x_i y_i$$

$$m \sum_{i=1}^{N} x_i + b \sum_{i=1}^{N} 1 = \sum_{i=1}^{N} y_i$$

To solve these, let's define

$$S_x = \sum_{i=1}^{N} x_i \qquad S_y = \sum_{i=1}^{N} y_i$$

$$S_{xy} = \sum_{i=1}^{N} x_i y_i \qquad S_{x^2} = \sum_{i=1}^{N} x_i^2$$

All of these come from our measured experimental data

$$m\, S_{x^2} + b\, S_x = S_{xy}$$
$$m\, S_x + b\, N = S_y$$

we want this vector

$$\begin{pmatrix} S_{x^2} & S_x \\ S_x & N \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} S_{xy} \\ S_y \end{pmatrix}$$

Using Cramer's rule:

$$m = \frac{1}{D}\left(N S_{xy} - S_x S_y\right)$$

$$b = \frac{1}{D}\left(S_x^2 S_y - S_x S_{xy}\right)$$

$$D = N S_{x^2} - (S_x)^2 \quad \longleftarrow \text{determinant}$$

The hardest part is finding a suitable linear model. For example, here's some data on the vapor pressure of water

T (°C)	P (torr)
0	4.579
5	6.543
10	9.209
15	12.788
20	17.535
25	23.756

heat of vaporization

The model : $p = c\, e^{-\Delta H_m / RT}$ ← not linear!

$$\ln p = \ln c - \frac{\Delta H_m}{RT}$$

So, if we let $y = \ln p$ and $x = \frac{1}{T}$ we'll have a linear model with

$$m = \frac{-\Delta H_m}{R} \qquad b = \ln c$$

$y = \ln p$	$x = 1/T$ (K^{-1})
1.521	0.00366
1.878	0.00360
2.220	0.00353
2.549	0.00347
2.864	0.00341
3.168	0.00335

$S_x = 0.0210$

$S_y = 14.200$

$S_{xy} = 0.0494$

$S_{x^2} = 7.3731 \times 10^{-5}$

→ $m = -5.3645 \times 10^{3}$

$b = 21.16355$

The correlation coefficient

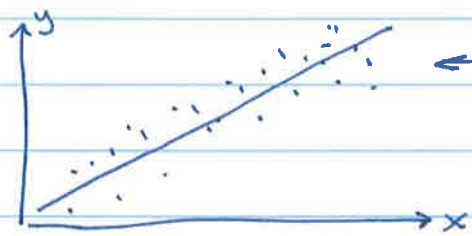
$$r = \frac{N S_{xy} - S_x S_y}{\sqrt{(N S_{x^2} - (S_x)^2)(N S_{y^2} - (S_y)^2)}}$$

If this is close to 1, we have a good fit

Note:

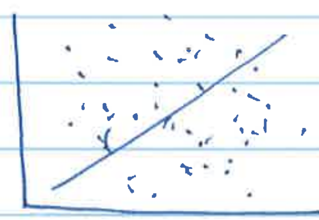
$$r = \frac{\text{Covariance}(x, y)}{\sigma_x \; \sigma_y}$$

$$\text{Covariance}(x, y) = \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})$$



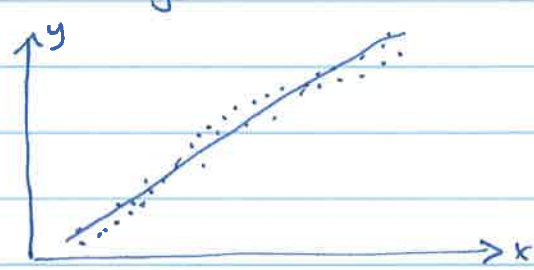
r tells you something about the width of the scatter around the best linear model

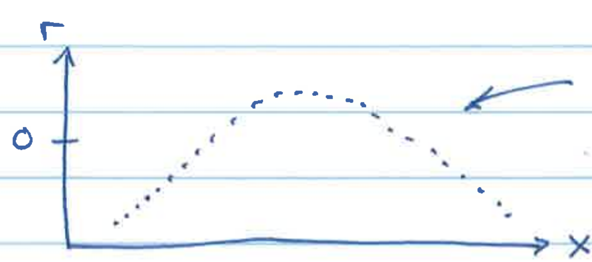
$r = 1 \rightarrow$ all points are on the line



$r = 0 \rightarrow$ no particular correlation

Plotting residuals can sometimes reveal flaws in your model





this structure means we're missing something parabolic in our model for the data