Big Numbers

- Avagadro's number is big

$$N_A = 6.02 \times 10^{23}$$

- But compare (log means natural log)

$$\log N_A \sim 50 \qquad (54.7 \text{ to be more exact})$$

- What about the number of rearrangements of the molecules in this room

$$N_A!$$

This is exponentially large. We will show in a sec, that

$$\boxed{N! \simeq \left(\frac{N}{e}\right)^N = N^N e^{-N} \leftarrow \text{Stirling Approx}}$$

Thus

$$\boxed{\log N! = N \log N - N \leftarrow \text{Stirling Approx}}$$

So

$$\log N_A! = N_A \underbrace{\log N_A}_{54.7} - N_A \simeq 53.7\, N_A = 3 \times 10^{25}$$

So even the $\log N!$ is a very large number. Lets call this exponentially large

Proof

$$N! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot N$$

$$\log N! = \log(1) + \log(2) + \log(3) + \ldots \log N$$

- This sum of logs can be replaced by an integral if N is large (see figure)

$$\log N! \simeq \int_1^N dx \log x$$

by parts

$$\simeq x \log x - x \Big|_1^N$$

$$\boxed{\log N! \simeq N \log N - N}$$

or since $N! = e^{\log N!}$

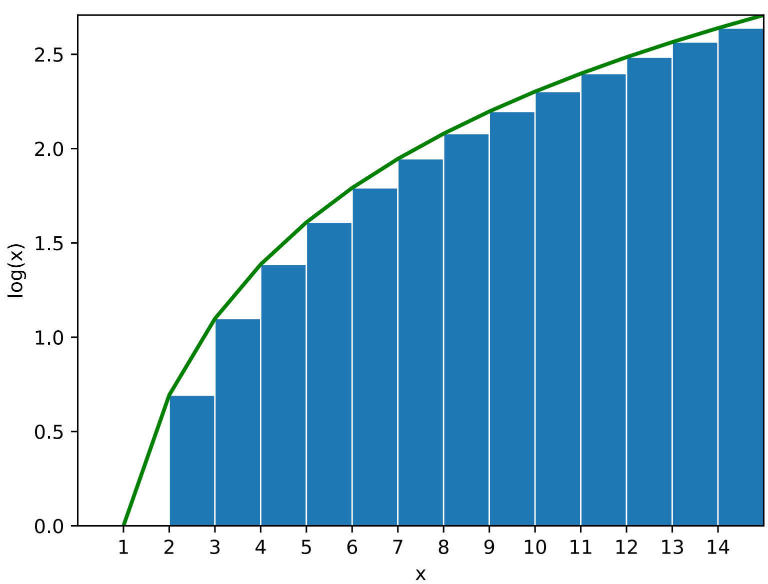
$$\boxed{N! \simeq N^N e^{-N}}$$

- You can find a better approximation, if you work harder (see book), which gives

$$N! = N^N e^{-N} \sqrt{2\pi N}$$

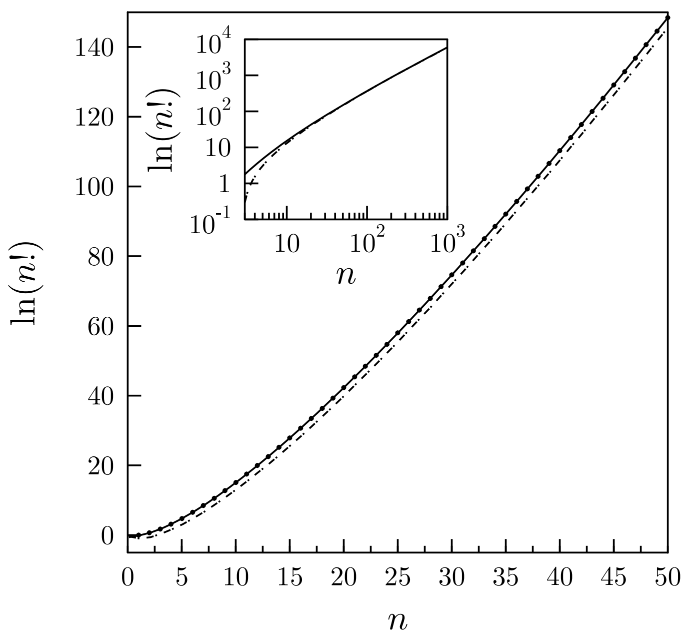
But we will not generally need the $2\pi N$.

Deriving the Stirling approximation:



Replace the sum with integral

Accuracy of Stirling



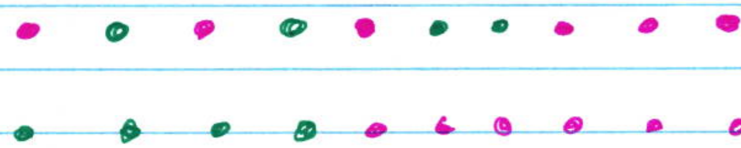
- Points: $\log(n!)$
- Dashed: $n \log n - n$
- Solid: $\log(n^n e^{-n} \sqrt{2\pi n})$

We will used the dashed

Combinatorics

- Consider the 10 atoms shown below. Four of them have been "activated"

- There are many ways to activate four some of which are shown below



$N = 10$ sites

$r = 4$ $\quad$ $N - r = 6$

- The total number ways of selecting 4 sites out of 10 is

$${}^{10}C_4 \equiv \frac{10!}{4!\,6!} \qquad \text{"10 choose 4"}$$

or selecting r out of N is

$${}^{N}C_r = \frac{N!}{r!\,(N-r)!} \longleftarrow \text{This is called a binomial coefficient}$$

That's because there are $N!$ rearrangements. But, rearrangements wich shuffle the green dots do not lead to a new selection. There are $r!$ of these. Similarly there $(N-r)!$ rearrangements of the pink dots. So the total number of possible selections is

$$^{N}C_{r} = \frac{N!}{r!\,(N-r)!}$$

Probability Distributions

- First imagine that a variable x takes a set of discrete outcomes x_i, e.g. a loaded dice with probability $\mathcal{P}_i$ $i = 1 \ldots N$, with $N = 6$ for loaded dice

Then
$$\sum_i \mathcal{P}_i = 1$$

$$\langle x \rangle = \sum_i x_i \mathcal{P}_i \qquad \text{also } \bar{x}$$

$$\langle x^2 \rangle = \sum_i x_i^2 \mathcal{P}_i \qquad \text{also } \overline{x^2}$$

The deviation from the mean is

$$\delta x \equiv x - \langle x \rangle$$

The mean deviation is

$$\langle \delta x \rangle = \langle x - \underbrace{\langle x \rangle}_{\text{const}} \rangle = \langle x \rangle - \langle x \rangle = 0$$

- The mean squared deviation, or variance

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

std. deviation squared

$$= \langle x^2 - 2x \underbrace{\langle x \rangle}_{\text{const}} + \underbrace{\langle x \rangle^2}_{\text{const}^2} \rangle$$

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

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

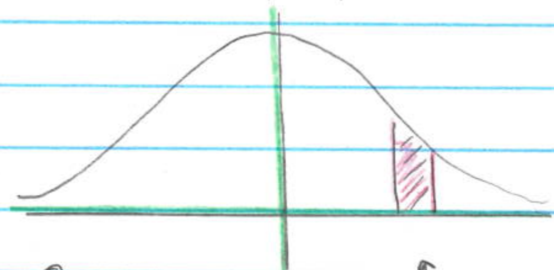
Continuous variables:

- The book (and other books) call it $P(x)$ for the probability density.

$$d\mathcal{P} = P(x)\, dx$$, probability to find x between x and $x + dx$

probabity

$P(x)$



x

$$\frac{d\mathcal{P}}{dx} = P(x)$$

$P(x)$ is the probability per x

Gaussian Distribution

$d\mathcal{P}$ = probability to be in this bin

Example: The Gaussian Distribution / Bell Curve

- $P(x) = N e^{-x^2/2\sigma^2}$
- Find N, $\langle x \rangle$, $\langle \delta x^2 \rangle$

Ok

$$1 = \sum d\mathcal{P} = \int P(x)\, dx$$

$$1 = \underbrace{\int_{-\infty}^{\infty} N e^{-x^2/2\sigma^2}\, dx}_{\text{Integral you should know}}$$

• If you dont know how to do this integral go read Appendix C2 through Eq. C6

$$1 = N \cdot (2\pi\sigma^2)^{1/2} \implies N = \frac{1}{\sqrt{2\pi\sigma^2}}$$

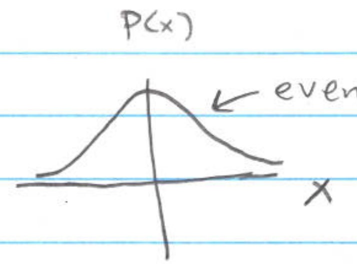
Thus we find the normalized Gaussian

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

note the units are 1/distance since σ has units distance

try to remember this

Now what about $\langle x \rangle$



$$\langle x \rangle = \int_{-\infty}^{\infty} x \; P(x) \, dx = 0$$

odd $\times$ even = odd

• Then we want $\langle x^2 \rangle$

$$\langle x^2 \rangle = \int_{-\infty}^{\infty} dx \; x^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2} = \sigma^2$$

There is a trick to doing this integral (and many like this) called a generating function which we will describe in homework.

Note that

$\langle x^2 \rangle$ has units (distance)^2 as it should.

Also

$$\sigma_x^2 = \langle x^2 \rangle - \cancel{\langle x \rangle}^{0\,2} = \sigma^2,$$ justifying the name.

Independent variables

- Consider random variable x, and random variable y, the probability that x is between $x+dx$ and y is between $y+dy$ is

$$d\mathcal{P} = P(x,y)\,dx\,dy$$

If x, and y are independent then the $P(x,y)$ factorizes

$$d\mathcal{P} = P_x(x)\,dx\,P_y(y)\,dy$$

Then

$$\langle xy \rangle = \int dx\,dy\,P(x,y)\;x\,y$$

$$= \int dx\,dy\,P_x(x)\,P_y(y)\,x\,y$$

$$= \int dx\,P_x(x)\,x \int dy\,P_y(y)\,dy\,y$$

$$= \langle x \rangle \langle y \rangle$$

Sums of independent variables

- Consider

$$Y = x_1 + x_2 + \dots\, x_n$$

where each x is drawn independently from the distribution $P(x)$

i.e $P(x_1, \ldots x_n) = P(x_1) P(x_2) \ldots P(x_n)\, dx_1\, dx_2 \ldots dx_n$

Then

$$\langle Y \rangle = \langle x_1 + \ldots + x_n \rangle$$

Clear n times the average of $\langle x \rangle$

$$\boxed{\langle Y \rangle = \langle x_1 \rangle + \langle x_2 \rangle + \ldots \langle x_n \rangle = n \langle x \rangle}$$

What about the variance of $\langle Y \rangle$?

$$\delta Y = Y - \langle Y \rangle$$

$$= (x_1 - \langle x_1 \rangle) + (x_2 - \langle x_2 \rangle) + \ldots x_n - \langle x_n \rangle$$

$$\delta Y = \delta x_1 + \delta x_2 + \ldots + \delta x_n$$

$$\langle \delta Y^2 \rangle = \langle (\delta x_1 + \delta x_2 + \ldots \delta x_n)^2 \rangle$$

$$= \langle \delta x_1^2 \rangle + \langle \delta x_2^2 \rangle + \ldots \langle \delta x_n^2 \rangle + \text{terms like } \langle \delta x_1 \delta x_2 \rangle$$

But the cross terms all vanish since

$$\langle \delta x_1 \delta x_2 \rangle = \langle \delta x_1 \rangle \langle \delta x_2 \rangle = 0$$

independence

So

$$\boxed{\sigma_Y^2 = \langle \delta Y^2 \rangle = n \langle x^2 \rangle = n\sigma_x^2}$$

- There is more that can be said. I will not prove it, but if n is large the probability of Y is gaussian, regardless of P(x)!

This is the Central Limit Theorem

$$\boxed{\begin{array}{l} P(Y) = \dfrac{1}{\sqrt{2\pi\sigma_Y^2}} e^{-(Y-\bar{Y})^2/2\sigma_Y^2} \\ \\ \text{with} \quad \sigma_Y^2 = n\,\sigma_x^2 \qquad \bar{Y} = n\langle x \rangle \end{array}}$$