Statistics

Needed to avoid having to deal with 10^{23} individual molecules and their degrees of freedom.

Discrete Distributions

E.g., flipping a coin has only 2 possible outcomes, each with $P_i = 1/2$ (for an unbiased toss)

Consider a function that depends on the outcome, e.g. I pay \$1 for every head, win \$2 for every tail.

Expectation value: $\langle f \rangle = \sum_i P_i f_i$

these angle brackets $\langle\rangle$ are called "ensemble average" in statistical mechanics

In the example, expectation of gain for a single coin toss is $\langle f \rangle = \frac{1}{2}(-1) + \frac{1}{2}(+2) = +0.5$

Properties: $\langle cx \rangle = c\langle x \rangle$ if c is constant

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

$\langle xy \rangle = \langle x \rangle \langle y \rangle$ if and only if x, y are independent!

Continuous Distributions

At limit of very many samples – e.g. height distr. of a large adult population, probability $P(h)$

$$\int_0^\infty P(h)\,dh = 1$$

$$\langle h \rangle = \int_0^\infty h P(h)\,dh$$ } mean height

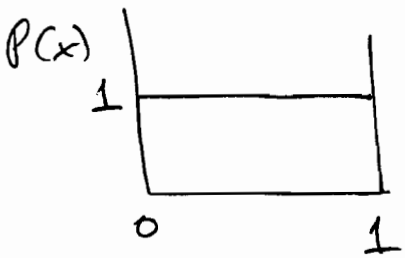
Variance $V(x) = \langle (x - \langle x \rangle)^2 \rangle$

Standard deviation $\sigma = \sqrt{V(x)}$

[for small samples, $\sigma = \sqrt{\frac{\sum (x - \langle x \rangle)^2}{n-1}}$ but diff. between $\sqrt{n-1}$ and $\sqrt{n}$ vanishes as $n \to \infty$]

Uniform Distribution

We frequently would like to sample "randomly" a given continuous variable, e.g., a torsional angle ← this requires sampling of a uniform distribution -



also can be used for discrete

- e.g. { if rand() < 0.5 "heads" else "tails" }

How does one generate uniformly distributed "random" numbers on a computer? (Discrete!)

Simple, efficient method: Linear congruential

$$x_{n+1} = (a x_n + c) \bmod (m)$$

↳ modulo - remainder of division

produces results between $\emptyset$ and $m-1$

e.g., $m = 2^4 = 16$ $a = 13$ $c = 1$

n	$\emptyset$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
x_n	10	3	8	9	6	15	4	5	2	11	$\emptyset$	1	14	7	12	13	10

↑ "seed"

period = 16 = m

→ Scale output rand() = x_n/m : (0,1)

Practical example "minimal standard" generator rand() in Press, "Numerical Recipes", 1992:

$a = 7^5 = 16807$ $\quad m = 2^{31} - 1 = 2147483647$ $\quad c = 0$

period $= m \approx 2.1 \cdot 10^9$

Limitations / problems: • low-order correlations (small seed X_0 gives small X_i)

• period exhaustion (many simulations $> 10^9$ steps)

Solutions: • "shuffle output" to reduce correlations

• combine w/ another for longer period, 10^{18}

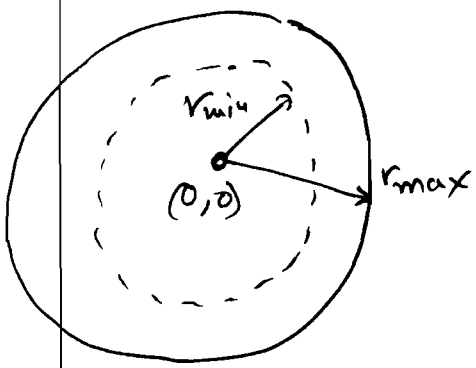
Other Distributions

Consider a diatomic molecule, e.g. N_2, in 2D

How do we generate a molecule in a "random" orientation, with "random" bond length between r_{min} and r_{max}? [Implies $U_{bond} = \begin{cases} 0 & r_{min} < r < r_{max} \\ \infty & \text{otherwise} \end{cases}$]

→ place first atom at (0,0)

→ where does the second atom go?



Slow, correct method:

→ $x_1 = (2 \cdot \text{rand}() - 1)\, r_{max}$

$y_1 = (2 \cdot \text{rand}() - 1) \cdot r_{max}$

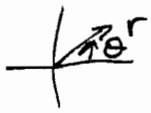
$r_1 = \sqrt{x_1^2 + y_1^2}$

if $r_1 < r_{max}$ and $r_1 > r_{min}$

accept

else reject

Fast, incorrect method: $r_i = r_{min} + rand() \cdot (r_{max} - r_{min})$

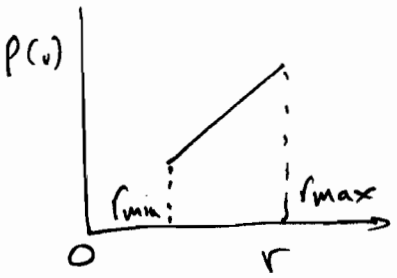
in polar coordinates (r, θ) 

! generate uniform length

$\theta = 2 \cdot \pi \cdot rand()$

$x_i = r_i \cos\theta \quad y_i = r_i \sin\theta$

Fast, correct method: take into account "Jacobian" of rectangular → polar transformation to correct for larger area at long r; need to sample

$P(r) \propto r$ 

→ 

Cummulative probability

$$H(r) = \int_{r_{min}}^{r} r\,dr \Big/ \int_{r_{min}}^{r_{max}} r\,dr \Rightarrow$$

$$H(r) = \frac{r^2 - r_{min}^2}{r_{max}^2 - r_{min}^2} = x \Rightarrow \quad H^{-1}(x) = \sqrt{x(r_{max}^2 - r_{min}^2) + r_{min}^2}$$

$$\therefore \quad r_i = \sqrt{rand()(r_{max}^2 - r_{min}^2) + r_{min}^2}$$

$$\theta = 2\pi\, rand()$$

$$x_i = r_i \cos\theta \quad y_i = r_i \sin\theta$$

Central Limit Theorem

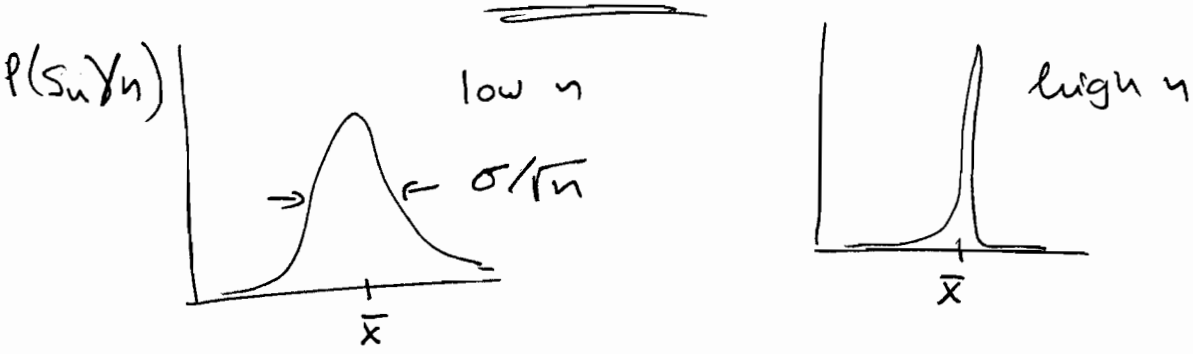
Applies to systems with many independent, random variables:

$x_1, x_2, \ldots x_n$ each with mean $\bar{x}$ and variance σ^2

The sample sum, $S_n = \sum_{i=1}^{n} x_i$ has expected mean $\langle S_n \rangle = n \cdot \bar{x}$

↑ average over many realizations of $x_1, x_2 .. x_n$

and follows a normal distribution (Gaussian) with variance σ^2/n



In other words, if you do the same, careful measurement n times, and the error in each measurement is m%, the error of n measurements is $m/\sqrt{n}$.