Basics of Probabilities & Statistics

History: Why did people need to come up with the theory of probability?

(1501-76) Cardano gives probabilities associated with dice

(1650) Chevalier de Mere - a gambler - asks Blaise Pascal why some bets make money while others don't.

Pascal & Fermat write Pierre de Fermat & together they found the theory of probability.

Basic Concepts:

What does 30% chance of rain mean?
What does it mean to say the chance of throwing a double 6 is 1/36?

Need an "ensemble" - collection of a large number of similarly prepared systems.

\[ P(\text{event } a) = \frac{\# \text{ of events resulting in } a}{\# \text{ of all possible events}} \]

Ex. Dice

One die: \( P(1) = \frac{1}{6} \)
\( P(2) = \frac{1}{6} \)
\( \vdots \)

Two dice: \( P_2(4) = \sum_{y} P_1(y) P_1(4-y) = \frac{3}{36} \)

Pictorially: \( \frac{3}{36} \)

# of ways of getting 4
/all possible combinations

How do we test this?

1. Conduct Exp. simultaneously with many systems.
2. Conduct Exp. over & over with same die.

Random Walks

Random Walks: Drunk takes steps length \( \ell \) either to the left or to the right.

\[ P(\text{Right}) = p \quad P= \theta = \text{Coin, etc} \]
\[ P(\text{Left}) = q \quad \text{in general } P \neq q \]
at all times \( x = ml \) where \( m \) is an integer

Q: After taking \( N \) steps what is probability of being at \( x = ml \)?

Boils down to adding \( N \) vectors of equal length but random direction

Important For:

1. **Polymer Configuration** - how big is a chain of atoms in solution important for manufacturing etc.
2. **Diffusion of a molecule or Colloidal particle** - how fast can you smell your dinner as he enters a room
3. **Magnetism** - what is magnetization of \( N \) randomly oriented spins
4. **Light diffusing through an opaque medium such as milk**

**DWS** Diffusive Wave Spectroscopy

**How do we Calculate** \( P_N(m) \)?

\[
N = N_1 + N_2 \\
M = N_1 - N_2 \\
m = 2N_1 - N
\]

\( N_1 \) = steps to the right
\( N_2 \) = steps to the left

is \( N \) is odd \( m \) is odd
is \( N \) is even \( m \) is even

is we have \( N \) statistically independent steps with probability \( p \) \& \( q \) any one sequence will have a probability, \( P \) \( \prod \) \( P_{P_1} \cdots P_{P_{N_1}} \) \( Q_{Q_1} \cdots Q_{Q_{N_2}} \)

**How many such sequences are there?**

Total \# of possible permutations for a given outcome:

\[
\frac{N!}{N_1!N_2!} \quad \text{doesn't count what you pick p or q several N1's}
\]

\( \text{can exchange all p's with one another} \)

\( \text{can exchange all q's with one another} \)

Q: What is the probability of taking \( N \) steps \( N_1 \) to the right \& \( N_2 \) to the left?

figure this out with your partner
\[ P_N(n) = \frac{N!}{n_1! n_2!} \binom{n_2}{n_1} \binom{N-n_2}{n_1} P^{n_1} (1-P)^{n_2} \]

related to Binomial Distribution
\[
(N+q)^N = \sum_{n=0}^{N} \frac{N!}{n_1! n_2!} p^{n_1} q^{n_2}
\]

For \( P_N(m) \) need 
\[ n_1 = \frac{1}{2} (N+m) \quad \text{and} \quad n_2 = \frac{1}{2} (N-m) \]

to sub

if \( p = q = \frac{1}{2} \)

\[ P_N(m) = \frac{N!}{(N+m)! (N-m)!} \left( \frac{1}{2} \right)^N 
\]

Peaked around \( m=0 \)

How do we know this? 
plug in some \#'s

\[ \text{if } m = N \quad \frac{N!}{N! (0)!} \left( \frac{1}{2} \right)^N \quad \text{if } N = 4 \quad \left( \frac{1}{2} \right)^4 \]

\[ \text{if } m = 0 \quad \frac{N!}{N! (N)!} \left( \frac{1}{2} \right)^N \quad 2^{24} \frac{1}{4} \left( \frac{1}{2} \right)^4 = 6 \left( \frac{1}{2} \right)^4 \]

**Mean Values**

Let's say we have a class that took a midterm. How do we calculate the average score?

1. Add all the scores and divide by the \# of students
2. \( \bar{x} \) = Score \cdot Fraction of students that got that score \( \bar{x} \) is almost always used this way

\[ \bar{x} = \sum_m x(m) P(m) \quad x(m) = \ell m \quad \bar{P}(m) = 1 \]

\[ \bar{f} = \sum_m f(m) P(m) \quad \bar{g} = \sum_m g(m) P(m) \]

\[ \bar{f(m)} + \bar{g(m)} = \sum_m P(m) [ f(m) + g(m) ] = \sum_m P(m) f(m) + \sum_m P(m) g(m) = \bar{f(m)} + \bar{g(m)} \]

\[ \bar{c \ f(m)} = c \ \bar{f(m)} \]

Let's calculate some things for random walks.
\[ \bar{n}_1 = \sum_{n_i=0}^{N} n_i \frac{p^n_i}{N!} \frac{q^{n-i}}{(n-i)!} \]

**Trick to be used over & over:**

\[ n_i p^{n_i} = p \frac{d}{dp} (p^{n_i}) \]

\[ \bar{n}_1 = \sum_{n_i=0}^{N} \frac{N!}{n_i!(n-n_i)!} p \frac{d}{dp} (p^{n_i}) q^{n-n_i} \]

Switch order of sum & derivative

\[ = p \frac{d}{dp} \left[ \sum_{n_i=0}^{N} \frac{N!}{n_i!(n-n_i)!} p^{n_i} q^{n-n_i} \right] \]

Why do we get to do this?

Partial amp goes through sum on \( n_i \)

\[ = p \frac{d}{dp} (p+q)^N = p N (p+q)^{N-1} = p N \]

---

**Dispersion about the mean**

**Second Central Moment**

\[ (\Delta n)^2 = \sum_{n_i=1}^{N} \bar{n}_1 (n_i - \bar{n}_1)^2 \geq 0 \]

\[ (n_i - \bar{n}_1)^2 = (n_i^2 - 2n_i \bar{n}_1 + \bar{n}_1^2) = n_i^2 - 2\bar{n}_1 n_i + \bar{n}_1^2 \]

\[ = \bar{n}_1^2 - n_i^2 \]

So we need to calculate \( \bar{n}_1^2 \) & \( n_i^2 \)

**Second Moment**

\[ \bar{n}_1^2 = \sum_{n_i=0}^{N} \frac{N!}{n_i!(n-n_i)!} p^{n_i} q^{n-n_i} n_i^2 \]

Use our trick above to show that

\[ \bar{n}_1^2 = n_i^2 + Npq \]

\[ (\Delta n)^2 = Npq \]

**Root Mean Square Deviation**

\[ = (\Delta n_1)^2 \]

\[ \text{a Standard Deviation} \]

This quantity tells us about the width of the distribution

\[ \frac{(\Delta n_1)^2}{n_1} = \frac{Npq}{Np} = \sqrt{\frac{1}{\frac{N}{n_1}}} \]

\[ \text{as } N \text{ distribution normal} \]
Central Limit Theorem

Outline main idea:
As \( N \to \infty \), distribution becomes smooth and continuous. Moreover,
\[ \ln \left( \frac{\text{histogram}}{\text{Gaussian}} \right) \to \bigwedge \]
Taking log of dist.
makes it even smoother. \[ \ln p \]

This allows us to use a power law expansion to approximate the distribution. Typically only the first five terms contribute.

Taylor expansion about \( x_0 \):
\[
f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \frac{f'''(x_0)}{3!}(x-x_0)^3
\]

If \( x_0 \) is a maximum,
\[
f(x_0) = \text{const}
\]
\[
f'(x_0) = 0
\]
\[
f''(x_0) = C_2 \text{ const} < 0
\]
\[
f'''(x_0) = C_3 \text{ small}
\]
\[
\ldots
\]

If \( (x-x_0) \) is small then we can keep only the lowest order term.
\[
\ln(p) = C_1 + \frac{1}{2} C_2 (x-x_0)^2 + \ldots
\]
\[
p = C_0 e^{\frac{1}{2} C_2 (x-x_0)^2} \text{ Gaussian distribution}
\]

This is the essence of the Central Limit Theorem.
Let \( X \) be a \( N \) i.i.d. normal random variable with parameters \( \mu = 0 \) and \( \sigma^2 = 9 \).

Then if \( \bar{X} \approx N \) \( \frac{X - \bar{X}}{\sigma / \sqrt{n}} \) has a standard normal distribution.

\[
\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i
\]

and

\[
\frac{X - \bar{X}}{\sigma / \sqrt{n}} \approx N(0, 1)
\]
Example: Colloidal particles in a microscope, minimum resolution ~ 0.1 μm

\[ x \ll \text{scale over which } p(x) \text{ changes} \]

\[
\sum_{m=0}^{N} p(m) = 1 \quad \Rightarrow \quad \int_{-\infty}^{\infty} p(x) \, dx = 1
\]

\( p(x) \) is the probability density. It is independent of \( dx \).

In order to get a probability, you need to multiply the probability density by an element of length:

\[
p(x) \, dx = \frac{1}{\sqrt{2\pi \sigma^2}} \, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx
\]

\[ \mu = \langle x \rangle \]
\[ \sigma = 2 \sqrt{\text{npq}} \]

\[ \mu = (p-q)N \ell \]

\[ \sigma = 2 \sqrt{npq} \ell \]

\[ \text{want to show} \quad \int_{-\infty}^{\infty} p(x) \, dx = 1 \]

\[ \text{want to calculate } \bar{x} \]

\[ \bar{x} = \int_{-\infty}^{\infty} x \, p(x) \, dx = \mu \]

\[ \text{want to calculate } \langle x^2 \rangle \]

\[ \langle x^2 \rangle = \int_{-\infty}^{\infty} (x-\mu)^2 \, p(x) \, dx \]

\[ \text{all of this requires integration of Gaussians} \]
Aside: Integrating Gaussians

$$I = \int_{-\infty}^{\infty} e^{-x^2} \, dx \quad I = \int_{-\infty}^{\infty} e^{-y^2} \, dy$$

$$I^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)} \, dx \, dy$$

Integrate over \( x-y \) plane

Switch to polar coordinates

\( x^2 + y^2 = r^2 \)
\( dx \, dy = r \, dr \, d\theta \)

$$I^2 = \int_0^{2\pi} \int_0^{\infty} e^{-r^2} \, r \, dr \, d\theta = 2\pi \int_0^{\infty} e^{-r^2} \, r \, dr$$

Trivial

$$I^2 = 2\pi \left. -\frac{1}{2} e^{-r^2} \right|_0^\infty = -\pi e^{-r^2} \bigg|_0^\infty = -\pi (0-1) = \pi$$

$$I = \sqrt{\pi}$$

$$\int_{-\infty}^{\infty} e^{-x^2} \, dx = \sqrt{\pi} \quad \text{or} \quad \int_{-\infty}^{\infty} e^{-x^2} \, dx = \sqrt{\pi}$$

**Conclusion:** \( I(n) = \int_{-\infty}^{\infty} e^{-ax^2} \, x^n \, dx \)

\( I(0) = \int_{-\infty}^{\infty} e^{-ax^2} \, dx = \frac{\sqrt{\pi}}{\sqrt{a}} \quad I(1) = \int_{-\infty}^{\infty} e^{-ax^2} \, x \, dx = \frac{\sqrt{\pi}}{2a} \)

What about higher moments? They can all be related through differentiation to \( I(0) \) & \( I(1) \) (See Appendix A4)

$$I(a) = -\frac{dI(0)}{da} = \frac{\sqrt{\pi}}{4a^{3/2}}$$

$$I(3) = -\frac{dI(1)}{da} = \frac{1}{2a^3}$$

If limits \( a \) is even, \( I(2) \) is finite and \( I(0) \) just multiply by \( 2 \) \(-a \)

If limits \( a \) is odd, \( I(2) \) is limits and odd \( I(2) \) over \( -a \) since \( \int_0^a \) integrating an odd function is zero.
Both to Probabilities

\[ \int_{-\infty}^{\infty} p(x) \, dx = \frac{1}{\sqrt{2\pi} \sigma} \int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2} \, dx \]

\[ = \frac{1}{\sqrt{2\pi} \sigma} \sqrt{\pi} \sqrt{2} \sigma = 1 \quad \checkmark \]

\[ \overline{x} = \int_{-\infty}^{\infty} x \, p(x) \, dx = \frac{1}{\sqrt{2\pi} \sigma} \int_{-\infty}^{\infty} x e^{-(x-\mu)^2/2\sigma^2} \, dx \]

\[ = \frac{1}{\sqrt{2\pi} \sigma} \left[ \int_{-\infty}^{\infty} ye^{-y^2/2\sigma^2} \, dy + \mu \int_{-\infty}^{\infty} e^{-y^2/2\sigma^2} \, dy \right] \]

\[ \checkmark \quad \overline{x} = \mu \frac{\sqrt{2\pi} \sigma}{\sqrt{2\pi} \sigma} = \mu \quad \overline{x} = (\mu - \mu) N \ell = \overline{m} \ell \]

\[ \Delta x^2 = (x-\mu)^2 = \int_{-\infty}^{\infty} (x-\mu)^2 \, p(x) \, dx \]

\[ = \frac{1}{\sqrt{2\pi} \sigma} \int_{-\infty}^{\infty} y^2 e^{-y^2/2\sigma^2} \, dy = \frac{1}{\sqrt{2\pi} \sigma} \left[ \frac{\sqrt{\pi} (2\sigma^2)^{3/2}}{\sigma} \right] \]

\[ \Delta x^2 = \sigma^2 \quad \checkmark \quad \boxed{\Delta x^2 = 4 NP \ell g \ell^2 = (\overline{m}^2 \ell^2)} \]

For a properly normalized Gaussian distribution, you can easily read off \( \overline{x} \), \( \sqrt{\Delta x^2} \)

\[ \text{Let's extend this methodology to more than one variable. Ex: Random walk in 2-D, 3-D} \]
**Discrete**

1-D \[ \sum P(m) = 1 \quad m \to x \quad \int p(x) \, dx = 1 \]

2-D \[ \sum \sum P(m, n) = 1 \quad m \to x \quad n \to y \quad \iint p(x, y) \, dx \, dy = 1 \]

is \( x \) \& \( y \) directions are not correlated

\[ \sum P(m) \leq P(n) = 1 \quad \int p(x) \, dx \leq \int p(y) \, dy = 1 \]

3-D \[ \sum \sum \quad P(m, n, e) = 1 \quad m \to x \quad n \to y \quad e \to z \quad \iiint p(x, y, z) = 1 \]

is \( x, y, z \) directions are uncorrelated

\[ \sum \sum P(m) \leq P(n) \leq P(e) = 1 \quad \int \int p(x) \, dx \, \int p(y) \, dy \, \int p(z) \, dz = 1 \]

**Averages**

Typically, we have some function \( F(x_i) \) \( i = 1, 2, 3 \ldots \) where average we are trying to evaluate

1-D \[ \overline{F} = \sum P(m) \cdot F(m) \quad \overline{F} = \int \overline{p(x)} \, dx \]

2-D \[ \overline{F} = \sum \sum P(m, n) \cdot F(m, n) \quad \overline{F} = \iint \overline{p(x,y)} \, dx \, dy \]

3-D \[ \vdots \]

4-D \[ \vdots \]

dm \ll dx \ll \text{Scale over which } F \text{ changes in } x \]

dn \ll dy \ll \text{Scale over which } F \text{ changes in } y
Summary

When approaching a new problem:

1. Determine probability $P$ or probability density $p$
   - Trick: Any distribution that dies off rapidly enough, in the limit of large $N$ approaches the Gaussian distribution (all we need to do is normalize, find $\mu$ and $\sigma$ deviation)

2. Calculate averages etc. by summing or integrating
   - For Gaussian distributions, $\mu$, $\sqrt{\sigma^2}$ can be read off immediately

3. Let's extend this methodology to systems of particles
Systems of Particles

Classical

- A State of System
  - faces of die showing

- Classical
  - \( \dot{q}_i = -\partial H / \partial p_i \)
  - Hamilton's eq's
  - energy is constant
  - \( \frac{q_i}{T_E} \):
  - trajectory

- Calculate # of ways of getting different results
  - \( P = \frac{\# of realizations}{\# of possible outcomes} \)

- Energy levels
  - Complicated systems with many degrees of freedom behave sufficiently "ergodically" that if we examine a system with energy \( E \) at a random time, the probability of finding it in region \( R \) of \( T_E \)
    - \( P[R] = \frac{\mathcal{V}[R]}{\mathcal{V}[T_E]} \)

- QM
  - QM
  - spin configuration
  - many HO need \( n_1, \ldots, n_N \)
  - HO in a box
  - vectors in a Hilbert space

- Ensemble
  - many realizations or ensembles
  - any of die faces are equally likely to show

- Basic Postulates
  - that need to be verified by experiment

- Tossing die each time (random)

- Schrödinger Eq.
  - \( \hbar \frac{\partial \psi}{\partial t} = -\frac{\partial H}{\partial t} \)
  - mixing of approximate states by small residual interactions between particles

- Subspace \( \mathcal{H}_E \) spanned by eigensystem of \( H \) with energy \( E \) values within some \( \Delta E \) of \( E \) where \( \Delta E < E \;

- Dimensions
  - dim \( \mathcal{H} \gg 1 \)
Ergodicity

In classical mechanics trajectories in space are determined by the Hamiltonian.

For each point \( x \in T \) define a Hamiltonian "vector field" \( \vec{h} \):

\[
\vec{h} = \left( \frac{\partial H}{\partial p_1}, \frac{\partial H}{\partial p_2}, \ldots, -\frac{\partial H}{\partial q_1}, -\frac{\partial H}{\partial q_2}, \ldots, -\frac{\partial H}{\partial q_n} \right)
\]

at each pt. we have a different value (magnitude & direction) for \( \vec{h} \).

Eq of motion \( \dot{x} = \vec{h} \) define dynamical trajectories or solutions to Hamiltonian & describe how system evolves.

What do this curve do?

So what do this curve do?

We know it stays within \( T_E \) unless Hamilton's equations do they wander around the whole surface?

Are the orbits closed?

Ex 1: 1-0 H.O.

Dimension of phase space is 2. Gorda \((q, p)\)

\[
H = \frac{1}{2} (p^2 + w^2 q^2)
\]

\[
\vec{h} = (p - w q)
\]

Ellipses (if \( w = 1 \) it's a circle)

Dimension of \( T_E = 1 \)
Ex 2 2 decoupled 1-d H.O.

dimension \( T^1 = 4 \) dim phase space \((q_1, q_2, p_1, p_2)\)

\[ H = H_1 + H_2 = \frac{1}{2} (p_1^2 + p_2^2 + \omega_1^2 q_1^2 + \omega_2^2 q_2^2) \]

Orbits all be on 4D surface of revolution about \( y \)

On the surfaces \( H_1 \) & \( H_2 \) are constant

family of ovals, one for \( H_1 \) & one for \( H_2 \)

Taurus

\[ \dim T_E = 2 \]

If ratio of frequencies is rational then orbits are closed. If ratio of frequencies is not rational then orbits are not closed & each orbit densely fill up the entire 2-torus.

The path gets arbitrarily close to any point we choose.

Note: We do not cover the space at all - Amount of area we cover is zero - Lines don't have any thickness

Ergodic Behavior

Expect complicated systems w/ many degrees of freedom (with suitably finite phase space \( T_E \) finite) to undergo some type of ergodic behavior in that dynamical orbits on \( T_E \) will "fill up" \( T_E \)

Let's discuss the possible types of ergodic behavior:

These are not necessarily true of real systems.
1. **Quasi Ergodic Behavior**

Almost all orbits "densely fill up" the whole volume $T_E$ except for a set of measure zero. Let $A_{y,\varepsilon} = \{ x \in T_E \mid \exists \text{ dynamical trajectory through } x \text{ fails to enter a ball of radius } \varepsilon \text{ around } y \}$. Then $V[A_{y,\varepsilon}] = 0$ for all $\varepsilon > 0$, all $y$.

2. **Metric Indecomposability**:

$T_E$ cannot be written as a disjoint union of subsets $A, B$ such that $V[A] > 0$, $V[B] > 0$.

$$A_\varepsilon = A, \quad B_\varepsilon = B$$

Orbits stay on their subsets and cannot divide all possible orbits into a number of subsets and add them up.

3. **Mixing Behavior**

(Sugar diffusing in water)

For any $A, B \subseteq T_E$, $\lim_{t \to \infty} V[A_t \cap B] = \frac{V[A]}{V[T_E]} = \frac{V[A]}{V[T_E]}$. A "uniformly" fills $T_E$ but still occupies the same volume $V[A]$ due to Liouville's Theorem.

4. **Strict Ergodic Behavior**

Every trajectory passes through every point on $T_E$.

This is what Boltzmann postulated.

This condition is impossible since $V[\text{traj}] = 0$.
Definitions are related to one another

3 implies 2: Since if A spreads over T_e uniformly then it must bleed into B as well

2 fails 3 fails: If we define sets that fail to satisfy 2 then A+B would also fail to satisfy 3

if not 2 then not 1: Since if we could separate A+B we could find a small \( \text{Ball of radius } \epsilon \) inside B such that a path starting at some point in A does not enter this ball

if not 1 then not 2: Since we could let \( A = A_y, \epsilon \) \( B = T_e - A_y, \epsilon \)

2 \( \iff \) 1 imply one another (I think this is true) but not sure

A: Does taurus example satisfy all three definitions?
(work this out in groups)

A: No, it violates definition #3 since if we pick a strip in phase space \& evolve it in time it moves uniformly without mixing.\[ \text{Solve } V[A \cap B] \neq V[A]V[B] \]

These three definitions give a flavor of what we mean by ergodic behavior. Research is still going on to determine what type of ergodicity is sufficient for making statistical mechanics work.
Birkhoff's Theorem

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t f(x_t') \, dt' = f(x)$$

(assumes metric indecomposability)

If the system traverses the phase space ergodically, then time average is equal to the space average.

In particular, let $f$ be some quantity associated with the system. Pressure, temperature, etc. Let coarse-grain the system.

Ex: a wall in a room filled with gas has some pressure in it. All the states of the gas that produce that pressure range can be grouped together.

Basic Hypothesis of Stat Mech:

Complicated systems with many degrees of freedom behave sufficiently "ergodically" that if we examine the system w/ energy $E$ at a random time the probability of finding it in region $R$ of $T_E$ is given by

$$P[R] = \frac{V[R]}{V[T_E]}$$

This probability distribution is called the "microcanonical ensemble."

Each point in $T_E$ is equally likely to be visited by any given trajectory through phase space.
Probability in Practice: The real reason stat mech works is not the collection of coarse grained observables \( \{F_1, \ldots, F_n\} \) that we measure but are about have the property that the region \( \mathcal{R} \subset \mathcal{R}_E \) where these observables take on their most probable values is such that \( V[\mathcal{R}] \approx V[\mathcal{R}_E] \).

The most probable region occupies nearly all of phase space. This is the notion of thermal equilibrium. Even if we start with the system in one of the improbable states, it will enter \( \mathcal{R} \) very quickly.

\[ F_i |_{\text{thermal equil}} = \text{The most probable value of } F_i(t) \approx \langle F_i \rangle \]

We could also speak about discrete states and count the number of states. We can divide \( \mathcal{R}_E \) into \( N \) dim cubes of size \( \hbar^{3N} \), where \( \hbar = \hbar \text{ Planck's constant} \) and momentum.

Cell size cannot be \( \leq \hbar^{3N} \) otherwise you violate the uncertainty principle. Cell size cannot be \( \geq \hbar^{3N} \) otherwise you lose information. Of course, other cell size will never enter into our calculations of physical quantities.