Broader picture: time scales + theories

Dynamical theory: ingredients
- a set of physical quantities $\tilde{y}$, multi-component vector
- given an initial distribution $P(\tilde{y}, 0) = P_0(\tilde{y})$
- a time evolution equation to calculate $P(\tilde{y}, t)$ at all $t \geq 0$
- additional parameters (time-independent) that encode info about the system

most generally:
- $n \rightarrow$ label of $n$th possible state in our system
- the theory is self-contained: no reference to other time-dependent quantities in the system
Stochastic processes + the laws of probability

system variable \( y(t) \) = labels the state of system at time \( t \)

discretize \( t_i = i \Delta t \), \( i = 0, 1, 2, \ldots \)

discretize \( y(t) \) \( \Rightarrow \) \( y_i \equiv y(i \Delta t) = n \)

where \( 1 \leq n \leq N \)

\[ \uparrow \]
total \# of States in System

trajectory = results of a single experiment, where states are recorded at time intervals \( \Delta t \) over some duration

\[ \text{trajectory} \equiv (y_0, y_1, y_2, \ldots, y_i) \]

ensemble = collection of trajectories from many repeated runs of the experiment

An ensemble is defined by how each experiment is prepared; the initial state \( y_0 \) is drawn from some prob. distribution \( P(y_0) \), where \( \sum_{y_0=1}^{N} P(y_0) = 1 \).
Two types:

If $P(y_0) = \delta_{y_0,n}$ ⇒ pure ensemble:

all experiments are initiated in same state $n$

Otherwise ⇒ mixed ensemble: the runs can have different starting states

Given an ensemble, the probability of a trajectory

$$P(\nu) = P(y_0, y_1, \ldots, y_i)$$

= # of trajectories with the state sequence $(y_0, y_1, \ldots, y_i)$

$$\sum_{\nu} P(\nu) = \sum_{y_0, y_1, \ldots, y_i=1}^N \text{traj}$$

sum over all possible trajectories

= 1

If a trajectory $\nu$ has an associated physical property $Q(\nu)$

⇒ ensemble average

$$\langle Q \rangle = \sum_{\nu} \frac{Q(\nu)}{\sum_{\nu} P(\nu)}$$
Basic notions of probability  \( \mathbb{P} \):

- let \( A, B \) be "events" drawn from an ensemble \( \mathcal{E} \), where an event is very general

- for example \( A = y_3 \) is an "event"
  or \( A = (y_0, y_1, y_2, y_3) \) is an "event", etc.

  \[ [A, B \text{ are quite general}] \]

joint probability \( \mathbb{P}(A, B) = \frac{\# \text{ trajectories where } A \text{ and } B \text{ occur}}{N_{\text{traj}}} \)

marginal probability \( \Rightarrow \) sum over subset of events

\[
\mathbb{P}(A) = \sum_B \mathbb{P}(A, B), \quad \mathbb{P}(B) = \sum_A \mathbb{P}(A, B)
\]

\[
\Rightarrow \quad \frac{\# \text{ traj. where } A \text{ occurs}}{N_{\text{traj}}}
\]

conditional probability: prob. of \( A \) given \( B \)

\[
\mathbb{P}(A \mid B) = \frac{\# \text{ traj. where } A \text{ and } B \text{ occur}}{\# \text{ traj. where } B \text{ occurs}} = \mathbb{P}(A, B) \]

\[
= \frac{\mathbb{P}(A, B)}{\mathbb{P}(B)}
\]

= prob. of \( A \) in a smaller ensemble of trajectories where \( B \) has to occur
\[
\sum_A \mathbb{P}(A|B) = \sum_A \frac{\mathbb{P}(A,B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1
\]

\[
\sum_B \mathbb{P}(A|B) \neq 1 \text{ in general}
\]

\[\Rightarrow \mathbb{P}(A|B) \text{ is normalized w/ respect to A, not B}\]

A and B are called independent if

\[\mathbb{P}(A,B) = \mathbb{P}(A)\mathbb{P}(B) \iff \mathbb{P}(A|B) = \mathbb{P}(A) \iff \mathbb{P}(B|A) = \mathbb{P}(B)\]

Remember conditionality is not causality:
A could be in the past or future of B
and \(\mathbb{P}(A|B)\) is still well-defined.

Crucial theorem relating reversal of conditionality:

\[
\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} \text{ Bayes' theorem}
\]

- follows trivially from definition \(\mathbb{P}(B|A) = \frac{\mathbb{P}(A,B)}{\mathbb{P}(A)}\)
- will be important in discussing time reversal later on
Three examples of Bayes theorem:

1) Disease testing: for example, rapid oral HIV test
   figure out: \[ P(\text{sick} | +\text{test}) = \frac{P(+\text{test} | \text{sick}) P(\text{sick})}{P(+\text{test})} \]
   what is known:
   \[ P(+\text{test} | \text{sick}) = 0.998 \quad \text{"sensitivity"} \]
   \[ P(+\text{test} | \text{not sick}) = 0.015 \quad \text{"prob. of false positives"} \]
   from demographics, \[ P(\text{sick}) = 0.003 \quad \text{in the U.S.} \]
   \[ P(\text{not sick}) = 0.997 \]
   \[ P(+\text{test}) = P(+) + P(+ | \text{not sick}) = P(+ | \text{sick}) P(\text{sick}) + P(+ | \text{not sick}) P(\text{not sick}) \]
   \[ = 0.018 \]
   \[ \Rightarrow P(+ | \text{sick}) = 0.17 \]

2) Monty Hall: game show:
   - car hidden behind 1 of 3 doors
   - goats behind the 2 others
   - contestant chooses one of the doors (door a)
   - before opening any, host randomly opens one of the doors you haven't chosen which he knows has a goat (door b)
   - question: asks contestant: do you want to switch your choice (door a \Rightarrow door c)

Does it make any difference?
\text{events}\\
A = \text{goat is behind door a} \quad \text{car is behind door b} \\
B = \text{C = car is behind door c} \\
O = \text{host opens door b} \quad [\text{hence } P(B|O) = 0] \\

\text{main question: } P(A|O) \equiv P(C|O) \\
\text{apply Bayes:} \\
P(A|O) = \frac{P(O|A) P(A)}{P(O)} \\
P(C|O) = \frac{P(O|C) P(C)}{P(O)} \\

\text{What we know:} \\
P(A) = P(B) = P(C) = \frac{1}{3} \\
P(O|A) = \frac{1}{2} \quad (\text{host chooses randomly between doors b and c}) \\
P(O|B) = 0 \quad (\text{host won't choose door with car}) \\
P(O|C) = 1 \quad (\text{host has only one choice of door with goat}) \\

P(O) = P(O|A) P(A) + P(O|B) P(B) + P(O|C) P(C) \\
= \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{3} \cdot 1 = \frac{1}{2} \\

\Rightarrow P(A|O) = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3} \\
P(C|O) = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}
So it always makes sense to switch
⇒ increases chance of winning!

3) Maximum likelihood estimation (MLE)

\[ p(M(\lambda) \mid D) = \frac{p(D \mid M(\lambda)) \cdot p(M(\lambda))}{p(D)} \]

\[ \Rightarrow \text{model that depends on unknown parameters} \ \lambda \]

\[ \Rightarrow \text{measured data} \]

\[ \Rightarrow \text{prob. that the model is with specific} \ \lambda \]

\[ \Rightarrow \text{is true, given the data} \]

\[ \Rightarrow \text{posterior} \]

Example: determining \( D \) from measurements of Brownian motion

\[ \lambda = \{ D \} \]

\[ M(\lambda) = \text{Einstein's theory for a particular value of} \ \lambda \]

\[ \text{data} \ D = \text{set of displacement measurements over large intervals} \ \Delta t \]

\[ \frac{1}{\sqrt{\pi D t}} e^{-\frac{x^2}{4 D t}} \]

\[ p(D(\xi_i, t_i) \mid \xi_i, \xi_{i+1}, \xi_{i+2}, \xi_{i+3}, \ldots) \]

\[ p(D(\xi_i, t_i) \mid \xi_i, t_i) \]

To estimate \( D \):

\[ \max_D p(M(\lambda) \mid D) = \max_D \frac{p(D \mid M(\lambda)) \cdot p(M(\lambda))}{p(D)} \]

\[ \Rightarrow \text{indep. of} \ D \]
Even when modeling measurements of non-probabilistic quantities (i.e. speed of light, mass of the Higgs)

\[ p(D|M) \] is still probabilistic because you should include apparatus error into your model.

\[ \Rightarrow \] In appropriate limits, MLE will recover standard statistics approaches, like minimizing \( \chi^2 \)

\[ \Rightarrow \] generally far better than naive curve-fitting or regression

For HW #3: use Bayesian analysis to estimate likelihood of extraterrestrial life.