Stochastic processes + the laws of probability

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

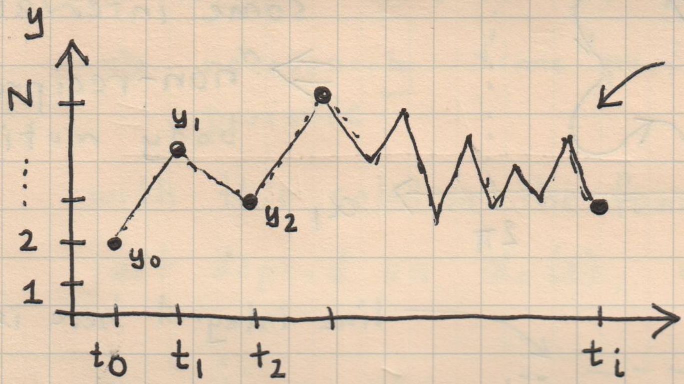
discretize $t_i \equiv i\,\delta t$, $i = 0, 1, 2, \ldots$

discretize $y(t) \Rightarrow y_i \equiv y(i\,\delta t) = n$

where $1 \le n \le N$

$\uparrow$ total # of states in system

trajectory $\equiv$ results of a single experiment, where states are recorded at time intervals δt over some duration



trajectory

$$\underbrace{(y_0, y_1, y_2, \ldots, y_i)}_{\equiv\, \nu}$$

$\equiv \nu$ shorthand notation for one trajectory

ensemble $\equiv$ collection of trajectories from many repeated runs of the experiment

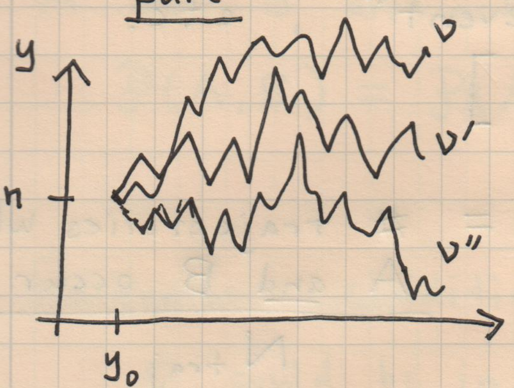
An ensemble is defined by how each experiment is prepared; the initial state y_0 is drawn from some prob. distribution $\mathcal{P}(y_0)$, where $\sum_{y_0=1}^{N} \mathcal{P}(y_0) = 1$.

Two types:

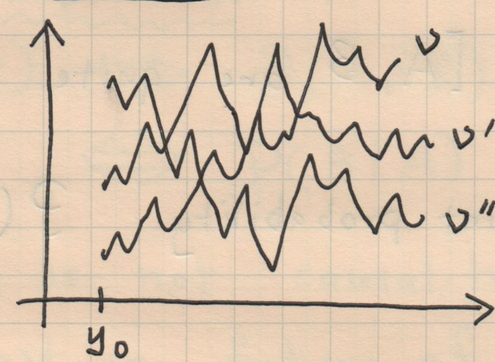
If $\mathcal{P}(y_0) = \delta_{y_0, n} \Rightarrow$ <u>pure ensemble</u>: all experiments are initiated in same state n

Otherwise $\Rightarrow$ <u>mixed ensemble</u>: the runs can have different starting states

<u>pure</u>



<u>mixed</u>



Given an ensemble, the probability of a trajectory

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

$$= \frac{\#\text{ of trajectories with the state sequence } (y_0, y_1, \ldots, y_i)}{N_{traj}}$$

total # of trajectories in ensemble (assumed to be arbitrarily large)

sum over all possible trajectories $\displaystyle\sum_\nu \mathcal{P}(\nu) = \sum_{y_0, y_1, \ldots y_i = 1}^{N} \mathcal{P}(y_0, y_1, \ldots y_N)$

$$= 1$$

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

$\Rightarrow$ ensemble average $\langle \overset{Q}{\cancel{A}} \rangle$ ~~or $\bar{A}$~~ $\displaystyle\equiv \sum_\nu \overset{Q}{\cancel{A}}(\nu)\, \mathcal{P}(\nu)$

Basic notions of probability $\mathcal{P}$:

- let A, B be "events" drawn from an ensemble ~~of~~, where an event is very general
- for example $A = y_3$ is an "event"

  <u>or</u> $A = (y_0, y_1, y_2, y_3)$ is an "event", etc.

[A, B are quite general]

joint probability $\mathcal{P}(A, B) = \dfrac{\#\text{ trajectories where A \underline{and} B occur}}{N_{traj}}$

marginal probability $\Rightarrow$ sum over subset of events

$$\mathcal{P}(A) = \sum_B \mathcal{P}(A, B) \, , \quad \mathcal{P}(B) = \sum_A \mathcal{P}(A, B)$$

$$\mathcal{P}(A) = \frac{\#\text{ traj. where A occurs}}{N_{traj}}$$

Conditional probability: prob. of A given B

$$\mathcal{P}(A \mid B) = \frac{\#\text{ traj. where A and B occur}}{\#\text{ traj. where B occurs}}$$

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

= prob. of A in a smaller ensemble of trajectories where B has to occur note: $\sum_A \mathcal{P}(A|B) = \sum_A \frac{\mathcal{P}(A,B)}{\mathcal{P}(B)} = \frac{\mathcal{P}(B)}{\mathcal{P}(B)} = 1$

$\sum_B \mathcal{P}(A|B) \neq 1$ in general

$\Rightarrow$ $\mathcal{P}(A|B)$ is normalized w/ respect to A, not B

A and B are called independent if

$$\mathcal{P}(A,B) = \mathcal{P}(A)\,\mathcal{P}(B) \iff \mathcal{P}(A|B) = \mathcal{P}(A)$$
$$\iff \mathcal{P}(B|A) = \mathcal{P}(B)$$

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

Crucial theorem relating reversal of conditionality:

$$\left.\mathcal{P}(A|B) = \frac{\mathcal{P}(B|A)\,\mathcal{P}(A)}{\mathcal{P}(B)}\right\}\text{ Bayes theorem}$$

- follows trivially from definition $\mathcal{P}(B|A) = \frac{\mathcal{P}(A,B)}{\mathcal{P}(A)}$
- will be important in discussing time reversal later on

Bayes Theorem Example

1) statement: "I have three kids, + at least one is a boy."

What is the probability that I have 3 boys?

$\tilde{B}$ = at least 1 is a boy

BBB = all three are boys

Answer:

all trajectories

GGG
GBG
BGG
GGB
GBB
BGB
BBG
BBB

$$P(BBB\,|\,\tilde{B}) = \frac{\#\text{ traj. w/ } BBB}{\#\text{ traj w/ } \tilde{B}} = \frac{1}{7}$$

Bayes: $P(BBB\,|\,\tilde{B})$

$$= \frac{P(\tilde{B}\,|\,BBB)\,P(BBB)}{P(\tilde{B})}$$

$$= \frac{1 \cdot \frac{1}{8}}{\frac{7}{8}} = \frac{1}{7} = 0.14$$

2) statement: "I have three kids. Each of them rolled a pair of dice yesterday. ~~At least~~ I have at least one boy who rolled snake eyes (1,1)."

What is the probability that I have three boys?

Did I learn any useful info to change my probability estimate?

Answer: $\epsilon = \frac{1}{36}$ = prob. to roll snake eyes

$\tilde{B}_s$ = at least 1 boy rolled snake eyes

Bayes theorem: $$P(BBB\,|\,\tilde{B}_s) = \frac{P(\tilde{B}_s\,|\,BBB)\,P(BBB)}{P(\tilde{B}_s)}$$

$$P(\tilde{B}_s \mid BBB) = 1 - \underbrace{(1-\epsilon)^3}_{\substack{\text{prob. none} \\ \text{of three} \\ \text{boys rolled snake eyes}}}$$

$$P(BBB) = \frac{1}{8}$$

$$P(\tilde{B}_s) = \sum_{X,Y,Z} P(\tilde{B}_s \mid XYZ)\, \overbrace{P(XYZ)}^{\frac{1}{8}}$$

$$= \frac{1}{8}\Big[\underbrace{P(\tilde{B}_s \mid GGG)}_{0} + \underbrace{P(\tilde{B}_s \mid BGG)}_{\epsilon} + \cdots + \underbrace{P(\tilde{B}_s \mid BBB)}_{1-(1-\epsilon)^3}\Big]$$

$$= \frac{1}{8}\left[3\epsilon + 3\left(1-(1-\epsilon)^2\right) + \left(1-(1-\epsilon)^3\right)\right]$$

$$= \frac{1}{8}\epsilon\left(12 - 6\epsilon + \epsilon^2\right)$$

Putting everything together: $P(BBB \mid \tilde{B}_s)$

$$= \frac{1-(1-\epsilon)^3}{\epsilon(12-6\epsilon+\epsilon^2)}$$

$$= 0.25 \quad \text{w/ } \epsilon = \frac{1}{36}$$

Higher probability!

Because playing dice + winning snake eyes actually conveyed something useful

$\Rightarrow$ rare event is more likely if I have more boys, increasing likelihood of BBB

Another interpretation: $\underbrace{P(BBB \mid \tilde{B}_s)}_{\substack{\text{"posterior" knowledge} \\ \text{about BBB after info} \\ \tilde{B}_s \text{ is given}}} = \underbrace{\left[\frac{P(\tilde{B}_s \mid BBB)}{P(\tilde{B}_s)}\right]}_{\text{conversion factor}} \underbrace{P(BBB)}_{\text{"prior" knowledge}}$

More useful application of Bayes:

Bayesian ~~maximum likelihood~~ model estimation

$M(\lambda)$ = physical model that depends on some unknown parameter(s) λ

D = measured data

$$P(M(\lambda)\,|\,D) = \left[\frac{\overbrace{P(D\,|\,M(\lambda))}^{\text{likelihood of } D \text{ given } M(\lambda)}}{P(D)}\right] P(M(\lambda))$$

$P(M(\lambda)\,|\,D)$: prob. that the model w/ a specific value of λ is true, given the data

"posterior"

$P(M(\lambda))$: "prior" encodes knowledge about possible models (values of λ) before taking data

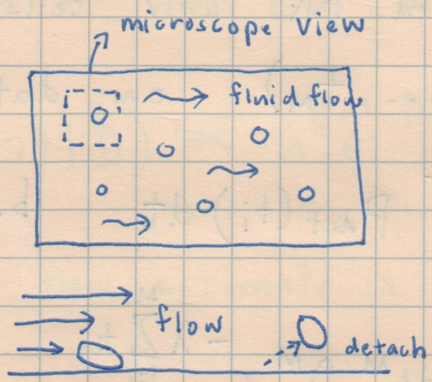
Goal: find most likely λ that fits the data

$\Rightarrow$ maximize right-hand side ~~term in brackets~~ w/ respect to all λ

$\Rightarrow$ maximize $P(D\,|\,M(\lambda))\,P(M(\lambda))$ b/c $P(D)$ independent of λ

called maximum a posteriori estimation (MAP)

Example



observe a population of attached cells in microscope under constant fluid flow

model: assume cells detach w/ probability per unit time λ

Estimate most likely value of λ from experimental measurements: record times of M detachment events $t_1, t_2, \ldots, t_M$

N_0 = initial # of attached cells $\gg 1$

$N(t)$ = # of attached cells at time t

$$N(t+dt) - N(t) \cong \underbrace{-\lambda N(t)\,dt}_{\substack{\text{\# of cells} \\ \text{detaching} \\ \text{b/t } t \text{ \& } t+dt}}$$ if # of cells is large

$$\Rightarrow \boxed{\frac{dN}{dt} = -\lambda N}$$

$$\Rightarrow N(t) = N_0 e^{-\lambda t}$$

$p_{det}(t)\,dt$ = prob. to ~~see a~~ see a particular cell detach between t & $t+dt$ (that was attached at $t=0$)

$$= \frac{\lambda N(t)\,dt}{N_0} = \lambda e^{-\lambda t}\,dt$$

Note: $\displaystyle\int_0^\infty p_{det}(t)\,dt = \int_0^\infty \lambda e^{-\lambda t}\,dt = -e^{-\lambda t}\Big|_0^\infty = 1$

If we sat & watched M indiv. cells detach,

$\mathcal{D} = (t_1, \ldots, t_M)$ our data

$$\mathcal{P}(\mathcal{D}\,|\,M(\lambda)) = \prod_{i=1}^{M} p_{det}(t_i)\,dt$$ b/c each cell is independent

prob. of data given a certain value of λ

$$= dt^M \lambda^M e^{-\lambda \sum_{i=1}^{M} t_i}$$

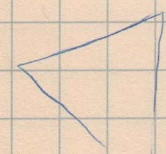
$\mathcal{P}(M(\lambda))$ = initially no clue so set constant $= C$

If $\mathcal{P}(M(\lambda))$ = constant $\Rightarrow$ maximize likelihood $\mathcal{P}(\mathcal{D}\,|\,M(\lambda))$

called maximum likelihood estimation (MLE)

$$\mathcal{P}(M(\lambda)\,|\,\mathcal{D}) = \frac{\mathcal{P}(D\,|\,M(\lambda))\,\mathcal{P}(M(\lambda))}{\mathcal{P}(D)}$$

$$\underset{\text{up to constant}}{\propto} \lambda^M e^{-\lambda \sum_i t_i}$$



maximize $\Rightarrow \frac{d}{d\lambda}\left(\lambda^M e^{-\lambda \sum_{i=1}^{M} t_i}\right) = 0$

$$\Rightarrow M\lambda^{M-1} e^{-\lambda \sum t_i} - \lambda^M e^{-\lambda \sum t_i} \sum t_i = 0$$

$$\Rightarrow \lambda^{M-1} e^{-\lambda \sum t_i}\left[M - \lambda \sum t_i\right] = 0$$

$$\Rightarrow \boxed{\lambda = \frac{M}{\sum_{i=1}^{M} t_i}}$$ best estimate

What if tomorrow you collect another L measurements t_j, $j = 1, \ldots, L$

Our prior probability of λ from previous experiment was $\propto \lambda^M e^{-\lambda \sum_{i=1}^{M} t_i}$

new $\mathcal{P}(M(\lambda)\,|\,\mathcal{D}) \propto \lambda^L e^{-\lambda \sum_{j=1}^{L} t_j} \cdot \lambda^M e^{-\lambda \sum_{i=1}^{M} t_i}$

maximize $\Rightarrow$ new estimate is $\lambda = \dfrac{M+L}{\sum_{i=1}^{M} t_i + \sum_{j=1}^{L} t_j}$

note this is <u>not</u> just an average b/t the estimates from each day's experiment