

Jointly Distributed Stochastic Variables

①

- $X_1, X_2, X_3, \dots, X_N$ are a set of jointly distributed stochastic variables if they are defined on the same sample space, S .

probability density: $P_{X_1 \dots X_N}(x_1, \dots, x_N)$

obeys

$$0 \leq P_{X_1 \dots X_N}(x_1, \dots, x_N) \leq 1$$

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} P_{X_1 \dots X_N}(x_1, \dots, x_N) dx_1 \dots dx_N = 1$$

- consider a pair of jointly distributed stochastic variables

$$P_{XY}(x, y)$$

- reduced probability density: $P_X(x) = \int dy P_{XY}(x, y)$

- n th moment of stochastic variable X

$$\langle x^n \rangle = \int dx dy P_{XY}(x, y) x^n = \int dx x^n P_X(x)$$

- joint moments of stochastic variables

$$\langle x^n y^m \rangle = \int dx dy x^n y^m P_{XY}(x, y)$$

- covariance: $\text{Cov}(X, Y) = \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle$
 $= \langle XY \rangle - \langle X \rangle \langle Y \rangle$

- correlation: $\text{Cor}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$

properties of correlation matrix

(2)

(i) $\text{Cor}(X, Y) = \text{Cor}(Y, X)$

(ii) $-1 \leq \text{Cor}(X, X) \leq 1$

(iii) $\text{Cor}(X, X) = 1$ $\text{Cor}(X, -X) = -1$

(iv) $\text{Cor}(aX+b, cY+d) = \text{Cor}(X, Y)$

- can extend to arbitrary number of stochastic variables

$$\text{Cor}_{ij} = \frac{\langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle}{\sigma_{x_i} \sigma_{x_j}} \quad \text{correlation matrix}$$

- for two independent variables X and Y

$$P_{XY}(x, y) = P_X(x) P_Y(y)$$

$$\Rightarrow \text{Cor}(X, Y) = 0$$

- the converse is not true, however: $\text{Cor}(X, Y) = 0 \not\Rightarrow P_{XY}(x, y) = P_X(x) P_Y(y)$

- several stochastic variables $\Rightarrow$ sometimes interested in stochastic variable which is a function of other stochastic variables

$$Z = G(X, Y) \quad \text{where } X, Y \text{ are stochastic variables}$$

$$P_Z(z) = \int_{-\infty}^{\infty} dx dy \delta(z - G(x, y)) P_{XY}(x, y)$$

example: $G(x, y) = x + y$

$$P_{XY}(x, y) = P_X(x) P_Y(y) \quad [X \text{ and } Y \text{ are independent}]$$

$$P_Z(z) = \int_{-\infty}^{\infty} dx dy \delta(z - x - y) P_X(x) P_Y(y)$$

$$= \int_{-\infty}^{\infty} dx P_X(x) P_Y(z - x) \Rightarrow \text{convolution}$$

$$G_Z(k) = G_X(k) G_Y(k)$$

- example of jointly distributed variables

(3)

random walk: a walker can move one unit to the left or the right, each ~~way~~ occurring with a probability of 50%. ~~then~~

starting at zero: $x_1 = \pm 1, x_2 = \pm 1, x_3 = \pm 1, \dots$

$x_1, x_2, x_3, \dots$ have the same sample space S

$$\langle (x_1 + x_2 + x_3 + \dots) \rangle = ?$$

for independent variables:

$$\langle x_1 + x_2 + \dots + x_N \rangle = \langle x_1 \rangle + \langle x_2 \rangle + \langle x_3 \rangle + \dots = 0$$

$$\langle (x_1 + x_2 + x_3 + \dots + x_N)^2 \rangle = \langle x_1^2 \rangle + \langle x_2^2 \rangle + \dots + \langle x_1 x_2 \rangle + \langle x_1 x_3 \rangle + \dots$$

for independent variables $\langle x_1 x_2 \rangle = \langle x_1 \rangle \langle x_2 \rangle = 0$

$$\Downarrow$$

$$\langle (x_1 + \dots + x_N)^2 \rangle = \sum_{i=1}^N \langle x_i^2 \rangle = N$$

$\Rightarrow$ distance travelled on average: zero

$\Rightarrow$ spread (average): proportional to the number of steps

$\Downarrow$
diffusion

- distribution of outcomes

for N steps: $Y = x_1 + \dots + x_N$

N - 1 way if all are $+1$

$N-2$ - N ways (one has to be -1
others $+1$)

$N-y$ $\frac{N!}{2^y (N-y)!}$ ways

$$P(N, y) = \frac{N!}{\left(\frac{N-y}{2}\right)! \left(\frac{N+y}{2}\right)!}$$

y -even number $\Rightarrow P(y) = \frac{N!}{\left(\frac{N+y}{2}\right)! \left(\frac{N-y}{2}\right)!}$

a note on the scientific method in the context of probability theory

④

- probability theory:

- a.) postulate an a priori probability distribution
- b.) perform appropriate mathematical transformation
- c.) compare a posteriori distribution with experimental observations

⇒ probability distribution cannot be observed, but can be reconstructed based on a large number of experiments

⇒ a priori probability distribution: depends on details of system

example: two dice

- a.) a priori the numbers 1 to 6 can occur with probability $1/6$ for each die
⇒ each outcome $1/36$ probability

- b.) transformation depends on question
for example ⇒ sum of results, distribution

$$P_z(z) = \int \delta(z - x - y) P_x(x) P_y(y) dx dy$$

discrete:

$$P_z(z) = \sum_{x,y} \delta_{z, x+y} P_x(x) P_y(y)$$

- c.) can then perform experiment to check whether die is OK

- statistical mechanics

⑤

- a.) assign equal probability to each microstate (a priori)
(ergodicity is necessary)
- b.) perform transformation (this depends on which ensemble is used)
 - predict interrelations between few ~~microscopic~~ macroscopic variables
- c.) check, based on experiment (or computer simulation)

BUT: in probability equal a priori probability depends on how a variable is defined

⇒ equal a priori probability corresponds to different physics if we take particle velocities from it we take some other variables to define a priori probabilities

scientific induction:

question: how can we deduce laws from a finite number of experimental results?

- from probability considerations: aim to assess that a hypothesised law is true, given a set of observations
- one can not answer this question unless an a priori probability of all possible hypotheses is assumed

for scientific theories ⇒ no 'a priori' theory of all hypotheses is available

→ agreement is based on corroborating experimental evidence for the a posteriori probability

Stochastic Processes

⑥

- consider a stochastic variable X
- One can use mappings to derive other stochastic variables from X

$$Y_X(t) = S(X, t) \text{ - stochastic process}$$

t - usually stands for time

- if a particular realization of X is considered

$$Y_x(t) = S(x, t) \rightarrow Y_x(t) \text{ is one function of } t$$

- for X (whole distribution) we have many realizations $\Rightarrow$ ensemble of realizations

- averages:

$$\langle Y(t) \rangle = \int Y_x(t) P_X(x) dx$$

- take time $t = t_1, t_2, \dots, t_n$

$$\langle Y(t_1) Y(t_2) \dots Y(t_n) \rangle = \int Y_x(t_1) Y_x(t_2) \dots P_X(x) dx$$

- example: autocorrelation function

$$\begin{aligned} K(t_1, t_2) &= \langle (Y(t_1) - \langle Y(t_1) \rangle) (Y(t_2) - \langle Y(t_2) \rangle) \rangle \\ &= \langle Y(t_1) Y(t_2) \rangle - \langle Y(t_1) \rangle \langle Y(t_2) \rangle \end{aligned}$$

$$\text{for } t_1 = t_2 \Rightarrow \sigma_Y^2 = K(t_1, t_1)$$

- stationary stochastic process:

$$\langle Y(t_1 + \tau) \dots Y(t_n + \tau) \rangle = \langle Y(t_1) \dots Y(t_n) \rangle$$

$\forall n, \tau$

- strictly stationary processes do not exist in nature/lab

- approximately realized if process itself lasts longer

than phenomenon under investigation

- autocorrelation or relaxation time τ_c :

time by which $K(t_1, t_1 + \tau)$ goes to zero

(time by which system forgets initial state)

$\Rightarrow$ if process in lab lasts longer than τ_c (much longer) ①
system can be assumed to be stationary

- stochastic quantity can have several components

$$K_{ij}(t_1, t_2) = \langle Y_i(t_1) Y_j(t_2) \rangle - \langle Y_i(t_1) \rangle \langle Y_j(t_2) \rangle$$

correlation matrix

- for zero-average stationary processes: $K(t_1, t_2) = K(|t_1 - t_2|)$

$\rightarrow$ one can always make stochastic process zero average

$$Y_i(t) \rightarrow Y_i(t) - \langle Y_i(t) \rangle$$

- example of stochastic process: Brownian motion (later!)

- for a stochastic process, we can write

$$P_1(y, t) = \int \delta(y - Y_x(t)) P_x(x) dx$$

the probability density at time t

one can also consider the joint probability density

$$P_n(y_1, t_1; y_2, t_2; \dots; y_n, t_n)$$

$$= \int \delta(y_1 - Y_x(t_1)) \delta(y_2 - Y_x(t_2)) \delta(y_3 - Y_x(t_3)) \dots P_x(x) dx$$

- calculating a distribution of the term

$\langle Y(t_1) \dots Y(t_n) \rangle$ is now a ~~joint~~ joint stochastic variables problem

$$\langle Y(t_1) \dots Y(t_n) \rangle = \int y_1 \dots y_n P_n(y_1, t_1; \dots; y_n, t_n) dy_1 \dots dy_n$$

- consistency conditions for quantity P_n

$$(i) P_n(y_1, t_1, \dots, y_n, t_n) \geq 0$$

(ii) P_n does not change when interchanging pairs
 $y_i, t_i \leftrightarrow y_j, t_j$

$$(iii) \int P_n(y_1, t_1; \dots; y_n, t_n) dy_n = P_{n-1}(y_1, t_1; \dots; y_{n-1}, t_{n-1})$$

$$(iv) \int dy_1 P_1(y_1, t_1) = 1$$

P_n - allow calculation of time averages $\Rightarrow$ complete specification of stochastic process

- usually P_n is the more accessible way of reconstructing a stochastic process

- conditional probability: $P_{n|1}(y_2, t_2 | y_1, t_1)$

- probability that at t_2 variable takes value y_2 given that at t_1 it took (or will take) y_1

$\Downarrow$
from the ensemble of realizations of the stochastic process $y_x(t)$ only those are considered which are at y_1 at t_1

$\Rightarrow$ subensemble

$$\int dy_2 P_{n|1}(y_2, t_2 | y_1, t_1) = 1$$

more generally: $P_{eln}(y_{n+1}, t_{n+1}, \dots, y_{n+l}, t_{n+l} | y_1, t_1, \dots, y_n, t_n)$

Bayes theorem:

$$\begin{aligned} P_{eln} &= P_{eln}(\dots) \\ &= \frac{P_{eln}(y_1, t_1, \dots, y_{n+l}, t_{n+l})}{P_n(y_1, t_1, \dots, y_n, t_n)} \end{aligned}$$

generalization of characteristic function to stochastic processes: characteristic functional

$$P_n(y_1, t_1, \dots, y_n, t_n) \rightarrow f(k_1, \dots, k_n) = \int e^{i(k_1 y_1 + \dots + k_n y_n)} P(y_1, t_1, \dots, y_n, t_n) dy_1 \dots dy_n$$

taking more and more sampling points (i.e. $n \rightarrow \infty$)

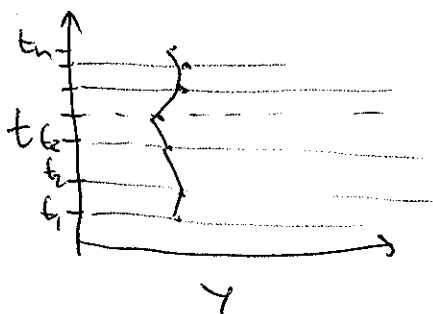
$$f(k_1, \dots, k_n) \rightarrow G[k] = \left\langle e^{i \int_{-\infty}^{\infty} k(t) y(t) dt} \right\rangle$$

notation $G(\{h\}) \Rightarrow$ depends on whole function $h(t)$

(9)

- we have also gone over into a path-integrated representation

$y, t_1, \dots, y_n, t_n \Rightarrow$ as $n \rightarrow \infty$ and the time intervals become closer to each other, these numbers specify a path



$$G(\{h\}) = \int D[y(t)] \tilde{P}[y(t)] e^{i \int h(t) y(t) dt}$$

$\tilde{P}[y(t)] \rightarrow$ probability of a particular path $y(t)$

the characteristic function can also be expanded as

$$G(\{h\}) = \sum_{m=0}^{\infty} \frac{i^m}{m!} \int h(t_1) \dots h(t_m) \langle y(t_1) \dots y(t_m) \rangle dt_1 \dots dt_m$$

$$\log G(\{h\}) = \sum_{m=0}^{\infty} \frac{i^m}{m!} \int h(t_1) \dots h(t_m) \underbrace{C_{1\dots m}(y)}_{\text{joint cumulant}} dt_1 \dots dt_m$$

for a stationary process:

$$P_n(y_1, t_1 + \tau, \dots, y_n, t_n + \tau) = P(y_1, t_1, \dots, y_n, t_n)$$

Gaussian process

$$G(\{h\}) = \exp \left[i \int h(t) \langle y(t) \rangle dt - \frac{1}{2} \int h(t_1) h(t_2) C_2(y_1, y_2) dt_1 dt_2 \right]$$

$\Rightarrow$ only first two cumulants contribute

often, but not always a reasonable approximation

Markov Processes

(10)

a Markov process is a stochastic process with

$$P_{11n-1}(y_n t_n | y_1 t_1; \dots; y_{n-1} t_{n-1}) = P_{111}(y_n t_n | y_{n-1} t_{n-1})$$

$\Downarrow$
conditional probability density of $y_n t_n$ depends only on $y_{n-1} t_{n-1}$, and not on data from earlier times

$\Downarrow$
Markov process uniquely determined by $P_1(y_1 t_1)$ and $P_{111}(y_2 t_2 | y_1 t_1)$ - (transition probability)

- we can write

$$\begin{aligned} P_3(y_1 t_1; y_2 t_2; y_3 t_3) &= \cancel{P_1(y_1 t_1)} \cancel{P_2(y_3 t_3; y_2 t_2 | y_1 t_1)} \\ &= P_2(y_1 t_1; y_2 t_2) P_{112}(y_3 t_3 | y_1 t_1; y_2 t_2) \\ &= P_1(y_1 t_1) P_{111}(y_2 t_2 | y_1 t_1) P_{111}(y_3 t_3 | y_2 t_2) \end{aligned}$$

- in this way any P_n can be written

$\Rightarrow$ manageable / useful in applications

- example: Brownian motion - large particle moving in a fluid of smaller particles

- can consider the velocity a Markov process

- if particle moves to the left $\Rightarrow$ collisions from the left are more likely

$\Rightarrow$ new velocity depends only on velocity at time immediately before collision

Chapman-Kolmogorov Equation

(11)

$$P_3(y_3, t_3; y_1, t_1; y_2, t_2) = P_1(y_1, t_1) P_{111}(y_2, t_2 | y_1, t_1) P_{111}(y_3, t_3 | y_2, t_2)$$

$$\int dy_2 P_3(y_3, t_3; y_1, t_1; y_2, t_2) = P_2(y_3, t_3; y_1, t_1) = P_1(y_1, t_1) P_{11}(y_3, t_3 | y_1, t_1)$$

$$\int dy_2 P_1(y_1, t_1) P_{111}(y_2, t_2 | y_1, t_1) P_{111}(y_3, t_3 | y_2, t_2)$$

$$\Rightarrow P_{111}(y_3, t_3 | y_1, t_1) = \int dy_2 P_{111}(y_3, t_3 | y_2, t_2) P(y_2, t_2 | y_1, t_1)$$

true for $t_1 < t_2 < t_3$

Markov Processes ~~obey~~ obey Chapman-Kolmogorov equation and the relation

$$P_1(y_2, t_2) = \int P_{111}(y_1, t_1 | y_2, t_2) P_1(y_1, t_1) dy_1$$

example: Wiener process

$$P_{111}(y_2, t_2 | y_1, t_1) = \frac{1}{\sqrt{2\pi(t_2 - t_1)}} \exp\left[-\frac{(y_2 - y_1)^2}{2(t_2 - t_1)}\right]$$

~~diffusion: $P_1(y, t) = P_1(y)$~~

with initial condition

$$P_1(y, 0) = \delta(y)$$

non-stationary Markov process

(for Brownian motion)

Stationary Markov Processes

(12)

- stochastic processes which are Markovian and stationary $\Rightarrow$ describe equilibrium fluctuations

$\Rightarrow P_i(y, t)$ is independent of time

$\Rightarrow P_i(y, t) = P_i(y)$ $P_i(y)$ is the equilibrium probability density

RC circuit: distribution of current



$$P(y_i) = \left(\frac{C}{2\pi kT} \right)^{1/2} \exp\left(-\beta \frac{C}{2} y_i^2\right)$$

$\Downarrow$
stationary Markov process

resistor with battery

stationary Markov process

~~$P_{ii}(y_1, t_2 | y_1, t_1)$~~ $P_{ii}(y_2, t_2 | y_1, t_1) = T_c(y_2 | y_1)$

depends only on $t_2 - t_1 = \tau$

$\Downarrow$
 $T_{c+n\tau}(y_2 | y_1) = \int dy_2 T_c(y_3 | y_2) T_c(y_2 | y_1)$

Chapman Kolmogorov equation