A Short Introduction to Probability Theory

Hagai Meirovitch

Probability is one of the basic concepts in science, appearing in quantum mechanics, statistical mechanics, and all branches of statistics.

It is essential for analyzing any scientific data.

Unfortunately, we shall be able only to touch upon very few points of this important topic.

Emphasize: Probability theory and computer simulation

Experimental probability

Rolling a die n times.

What is the chance to get an odd number?

n	10	50	100	400	1000	10,000 ….
m	7	29	46	207	504	5,036 ….

Relative frequency: $f(n)=m/n$

0.7 0.58 0.46 0.517 0.5040 0.5036 ….

$f(n) \rightarrow P = 0.5$; $P =$ experimental probability

While the exact P can never be obtained experimentally its notion and related properties are of great interest. To treat them efficiently we describe them within the framework of an *idealized* mathematical model - probability space.

Sample space

Elementary event (a possible outcome).

Tossing a coin – A happened, or B happened.

Rolling a die - 1, 2, 3, 4, 5, or 6 happened.

Event: any combination of elementary events.

An even number happened → one of the following happened: (2,4,6); a number larger than 3 happened (4,5, or 6).

The empty event: impossible event – ∅

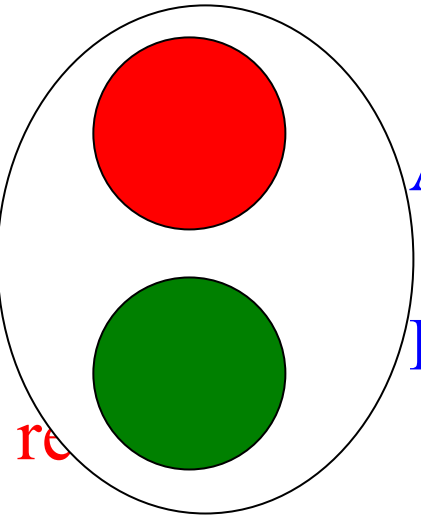
(2 < a number < 3).

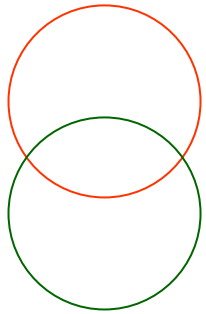
The certain event – Ω (Coin: A or B happened).

Complementary event - $\bar{A}=\Omega-A$ (1,2,3,4,5,6) - (2,4,6) = (1,3,5).

Union – (1,2,3) ∪ (2,3,5) = (1,2,3,5).

Intersection – (1,2,4) ∩ (2,5,6) = (2) (common)



A ⇗ B = 

A ⇗ B = intersection

+green

A ⇖ B = A and B

A ⇖ B = whole

Elementary probability space

- The sample space consists of a finite number n of points (elementary events) B.
- Every partial set is an event.
- A probability P(A) is defined for each event A.
- The probability of an elementary event B is $P(B)=1/n$.
- $P(A)=m/n$; m- # of points in event A.

Properties of *P*

- $0 \le P(A) \le 1$
- $P(A \cup B) \le P(A) + P(B)$
- $\sum_i P(A_i) = 1$ (A_i, elementary events).

Examples: a symmetric coin; an exact die.

However, in the experimental world a die is not exact and its rolling is not random; thus, the probabilities of the elementary events are not equal.
On the other hand, the probability space constitutes an ideal model with equal *P's*.
Comment: In general, elementary events can have different probabilities (e.g., a non-symmetric coin).

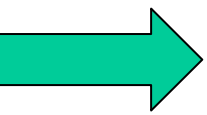
Example:

A box contains 20 marbles, 9 white, 11 black. A marble is drawn at random.

What is the probability that it is white?

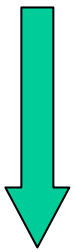
Elementary event (EE): selection of one of the 20 marbles.

Probability of EE: 1/20
The event A – a white marble was chosen contains 9 EE,



P=9/20 This consideration involves the ideal probability space; the real world significance: P is a result of many experiments, $P=f(n)$, n✹◎.

A problem related to to the experimental world (solution requires carrying out experiments: probability from statistics)



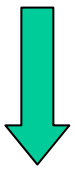
modeling
(idealization)

analytical solution difficult

Probability Space

Simulations in prob. space

(computer experiments: probability (and other properties) from statistics



Answer by analytical techniques, if possible

More complicated examples:

What is the number of ways to arrange r different balls in n cells?

Every cell can contain any number of balls.
Each ball can be put in n cells ⇓

$$\#\text{ ways} = n \cdot n \cdot n \cdot \ldots n = n^r$$

n^r= the number of words of length r (with repetitions) based on n letters. A,B,C ⇓ AA, BB, AB, BA, CC, CA, AC, BC, CB

$= 3^2 = 9$

Permutations

of samples of r objects out of n objects without repetitions (the order is considered) ($n!=1\cdot 2\cdot 3\cdot \ldots \cdot n$, $0!=1$):

$$(n)_r = n(n-1)\ldots(n-r+1) = \frac{1\cdot 2 \ \ldots \ (n-r)(n-r+1)\ldots n}{1\cdot 2 \ \ldots (n-r)} = \frac{n!}{(n-r)!}$$

$$(n)_n = n!$$

$(3)_2$ 🕒 (1,2), (1,3), (2,1), (2,3), (3,1), (3,2) (1,3) 🕒 (3,1)

of r (2) letter words from n (3) letters: AB, BA, CA, AC, BC, CB

= 6

Problem:

Find the probability that r people ($r \le 365$) selected at random will have different birthdays?

Sample space: all arrangements of r objects (people) in 365 cells (days) (EE) – their number = 365^r $\Rightarrow$ $p(\text{EE}) = 1/365^r$

Event A: not two birthdays fall on the same day-

# of points in A = $365 \cdot 364 \cdot \ldots \cdot (365-r+1) = (n)_r$

$$P(\text{A}) = \frac{(n)_r}{365^r} = \frac{365!}{(365-r)!365^r}$$

Combinations

of ways to select r out of n objects if the order is not considered.

# of combinations = # of permutations / $r!$

$$\binom{n}{r} = \frac{(n)_r}{r!} = \frac{n!}{(n-r)!r!} \qquad \binom{n}{0} = 1$$

$n=3;\ r=2$

Permutations/2

(1,2) (2,1) /2 (1,3) (3,1) /2 (2,3) (3,2) /2

Problem: In how many ways can n objects be divided into k groups of $r_1, r_2, \ldots..r_k$; $\swarrow$ $r_k = n$ without considering the order in each group but considering the order between the groups?

$$\binom{n}{r_1} \bullet \binom{n-r_1}{r_2} \bullet \binom{n-r_1-r_2}{r_3} \ldots\ldots = \frac{n!}{r_1! r_2! \ldots r_k!}$$

Multinomial coefficient: $(x_1 + x_2 + x_3 \ldots + x_k)^n$ $\quad x_1^{r1} x_2^{r2} x_3^{r3} \ldots x_k^{rk}$

Problem: How many events are defined in a sample space of n elementary events?

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \ldots.. + \binom{n}{n} = 2^n$$

$\binom{n}{r}$ Binomial coefficient

$$(1+x)^n = \binom{n}{0} + \binom{n}{1} x + \ldots \binom{n}{r} x^r + \ldots.. + \binom{n}{n} x^n$$

Problem:

23 chess players are divided into 3 groups of 8, 8, and 7 players. What is the probability that players A, B, and C are in the same group (event A)?

EE – an arrangement of the players in 3 groups.

EE = 23!/(8!8!7!)

If A, B, and C in the first group the # of arrangs. 20!/(5!8!7!)

etc.

$$P(\mathrm{A}) = \frac{20!\cdot 2}{5!8!7!} + \frac{20!}{8!8!4!} \Big/ \frac{23!}{8!8!7!} = \frac{21}{253}$$

Problem:

What is the number of ways to arrange n objects that $r_1, r_2, r_3, \ldots r_k$ of them are identical, $\swarrow r_i = n$?

If all are different, the number of permutations is $n!$ because $r_1, r_2, \ldots r_n$ are the same one has to divide $n!$ by $r_1!\, r_2! \ldots r_n!$

$$\#\text{ ways} = \frac{n!}{r_1!\, r_2! \ldots r_k!}$$

$\underline{6}\ \ 6\ \ \underline{24}$

55511122 22 one permutation

51152512 22 another permutation

Problem: 52 cards are divided among 4 players. What is the probability that every player will have a king?

EE – a possible division of the cards to 4 groups of 13.

EE $52!/(13!)^4$ (p.14)

If every player has a king, only 48 cards remained to be distributed into 4 groups of 12 ➧ # of EE(A) = $48!/(12!)^4$

$$P(A)=[4!48!/(12!)^4]/[52!/(13!)^4]$$

Product Spaces

So far the probability space modeled a single experiment – tossing a coin, rolling a die, etc. In the case of n experiments we define a product space:

Coin; two EE: 0 ,1
symmetric - $P=1/2$

$$\binom{1}{1,0} \times \binom{2}{1,0} \times \binom{3}{1,0} \times \ldots \binom{n}{1,0}$$

1 2 ……. n 1 ……. n

EE (vector): (1,0,0,1,…,1); (0,1,1,1,…,0); (…….); # (EE)=2^n

If the experiments are independent, $P(\text{EE})=(1/2)^n$

Die: EE are: 1, 2 , 3 , 4 , 5 ,6

$$\begin{matrix}1\\ \begin{pmatrix}1\\2\\3\\4\\5\\6\end{pmatrix}\end{matrix} \times \begin{matrix}2\\ \begin{pmatrix}1\\2\\3\\4\\5\\6\end{pmatrix}\end{matrix} \times \begin{matrix}3\\ \begin{pmatrix}1\\2\\3\\4\\5\\6\end{pmatrix}\end{matrix} \times \ldots \times \begin{matrix}n\\ \begin{pmatrix}1\\2\\3\\4\\5\\6\end{pmatrix}\end{matrix}$$

1……..n 1……..n

EE= (1,5,2,4….,3,6); (2,3,2,5,….,1,1); (4,…….)…; #(EE)=6^n

In the cases of independent experiments, $P(EE) = (1/6)^n$

Problem:

15 dice are rolled. Find the probability to obtain three times the numbers 1,2, and 3 and twice, 4, 5, and 6?

EE: all possible outcomes of 15 experiments, #(EE)= 6^{15}

#(A): according to formula on p. 16: $15!/[(3!)^3(2!)^3]$

$$P(\mathrm{A}) = \frac{15!}{6^{15}(3!)^3(2!)^3}$$

Also: 1 can be chosen in $(15 \prec 14 \prec 13)/3!$ ways.

2 in $(12 \prec 11 \prec 10)\ )/3!$ ways etc.

Dependent and independent events

Event A is independent of B if $P(A)$ is not affected if B occurred.

$$P(A/B)=P(A)$$

$P(A/B)$ – conditional probability. For example, die:

Independent:

$P(\text{even})=1/2$; $P(\text{even/square})=1/2$ [$P(\{2,4,6\}/\{1,4\})=1/2$]

Dependent:

$P(2)=1/6$; $P(2/\text{even})=1/3$

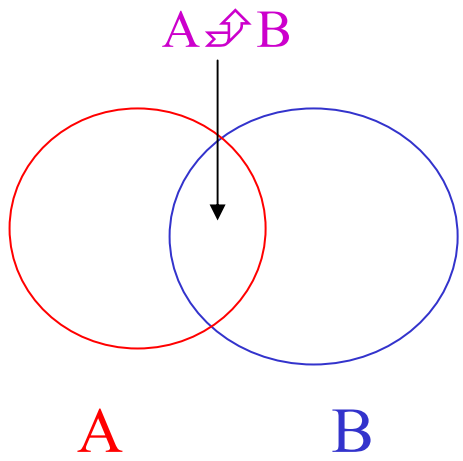
$P(\text{even/odd})=0$, while $P(\text{even})=1/2$ (disjoint)

Bayes Formula

$P(A/B) = P(A \cap B)/P(B)$; $P(B/A) = P(A \cap B)/P(A)$;

$P(A)>0$; $P(B)>0$

$$P(A/B) = \frac{P(B/A)\cdot P(A)}{P(B)}$$



A=(2) B=even (2,4,6) ⇓ $P(A/B)=1/3$. Using formula (line 2): $P(B)=1/2$; $P(A \cap B)=1/6$ ⇓

$$P(A/B) = \frac{1/6}{1/2} = 1/3$$

Independency: $P(A \cap B) = P(A)P(B)$

$1/6 \neq 1/6 \times 1/2$

If an event A must result in mutually exclusive events, $A_1,\ldots A_n$, i.e., $A=A\cap A_1 + A\cap A_2 +\ldots A\cap A_n$ then

$$P(A)=P(A_1)P(A/A_1) +\ldots\ldots+P(A_n)P(A/A_n)$$

Problem: Two cards are drawn successively from a deck. What is the probability that both are red?

EE- product space: (r,r), (r,no), (no,r), (no,no)

A – first card is red {(r,r); (r,no)}

B – second card is red {(no,r); (r,r)}

$$P(A\cap B)=P(r,r)=P(A)P(B/A)=1/2\cdot(25/51)$$

Summary

We have defined experimental probability as a limit of relative frequency, and then defined an elementary probability space, where P is known exactly. This space enables addressing and solving complicated problems without the need to carry out experiments.

We have defined permutations, combinations, product spaces, and conditional probability and described a systematic approach for solving problems.

We shall now define another important tool- a function called random variable, which enriches significantly the usefulness of probability spaces.

Random variables

For a given probability space with a set of elementary events $\{\blacklozenge\}=\Phi$, a random variable is a function $X=X(\blacklozenge)$ on the real line $-\infty < X(\blacklozenge) < \infty$.

Examples:

Coin: p - head; q- tail. One can define $X(p)=1$; $X(q)=0$. However, any other definition is acceptable - $X(p)=15$ $X(q)=2$, etc., where the choice is dictated by the problem.

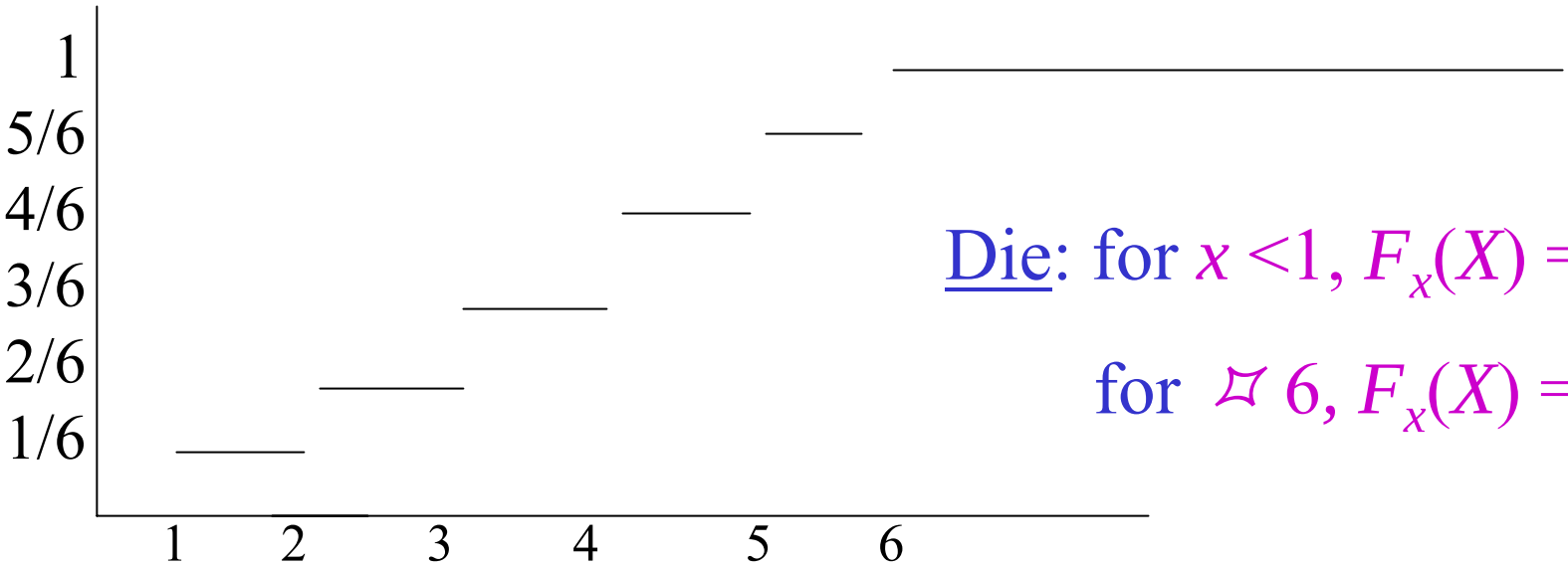
Tossing a coin n times, the sample space is vectors (1,0,0,1…) with $P(1,0,0,1\ldots)$. One can define $X=m$, where m is the number of successes (heads) (not an one-to-one function!)

Distribution Function (DF) or Cumulative DF

For a random variable X $(-\infty < X(\omega) < \infty)$

$$F_x(X) = P[X(\omega) \le x]$$

$F_x(X)$ is a monotonically increasing function



Die: for $x < 1$, $F_x(X) = 0$

for ≥ 6, $F_x(X) = 1$

Random variable of Poisson

$$P(x=m)=\frac{\lambda^m}{m!}\exp(-\lambda)$$

m=0,1,2,3,…

$$F(x)=\sum_{m\le x}\frac{\lambda^m}{m!}\exp(-\lambda)$$

So far we have discussed discrete random variables.

Continuous random variable – if $F(x)$ continuous and its derivative $f(x) = F'(x)$ is also continuous. $f(x)$ is called the probability density function; $f(x)dx$ is the probability between x and $x+dx$.

$$F(x)=\int_{-\infty}^{x} f(t)dt;\quad \int_{-\infty}^{\infty} f(t)dt=F(\infty)=1$$

The normal random variable

$$f(x)=\frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2}) \qquad \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp(-\frac{t^2}{2})dt=1$$

$$P(X\le x)=F_x(X)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp(-\frac{t^2}{2})dt$$

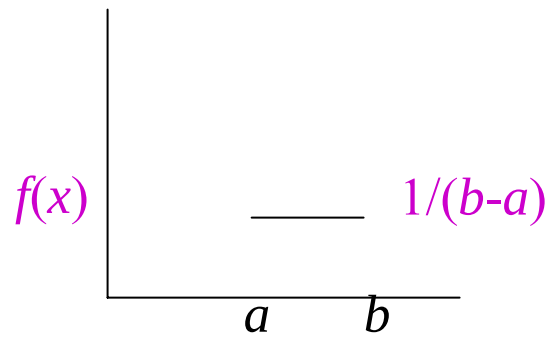
The uniform random variable

$$f(x)=\begin{cases}\text{const.} & a\le x\le b\\ 0 & \text{otherwise}\end{cases}$$

f(*x*) is constant=*c* between *a* and *b*

$$1=\int_{-\infty}^{+\infty}f(x)=c\int_{a}^{b}dx=c(b-a)\Rightarrow f(x)=\frac{1}{(b-a)}$$

$$F_x(X)=\int_{-\infty}^{x}f(t)dt=\begin{cases}0 & x<a\\ \frac{x-a}{b-a} & a\le x\le b\\ 1 & x\ge b\end{cases}$$



Expectation Value

X is a discrete random variable with n values, $x_1, x_2, \ldots x_n$. and $P(x_1), P(x_2), \ldots P(x_n)$, the expectation value $E(X)$ is:

$$E(X) = \mu = \sum_{i=1}^{n} P(x_i) x_i$$

Other names are: mean, statistical average.

Coin:

X 🕓 1 ; 0 with P and 1-P.

$E(X) = P \lessdot 1 + (1-P) \lessdot 0 = P$

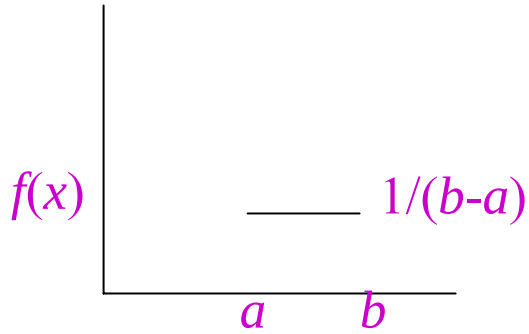
Die: X 🕓 1, 2, 3, 4, 5, 6 with P =1/6 for all.

$E(X) = (1/6) \lessdot (1+2+3+4+5+6) = 21/6 = 3.5$

Continuous random variable with $f(x)$

$$E(X)=\mu=\int_{-\infty}^{\infty} xf(x)dx$$

Provided that the integral converges.



Uniform random variable

$$E(X)=\int_a^b x\frac{1}{b-a}dx=\frac{x^2}{2(b-a)}\Bigg|_a^b=\frac{b^2-a^2}{2(b-a)}=\frac{b+a}{2}$$

$E(X)$ is a linear operator: If X & Y are defined on same space

$$E(X+Y)=E(X)+E(Y)\ ;\ \ E(CX)=CE(X)$$

C= const.

♦[X(♦)+Y(♦)]P(♦) = ♦ X(♦)P(♦) + ♦ Y(♦)P(♦) ♦

Variance

$$V(X) \equiv \sigma^2(X) = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

$$V(X) = E[(X-\mu)^2] = E[X^2 + \mu^2 - 2X\mu] =$$

$$= E(X^2) + \mu^2 - 2\mu E(X) = E(X^2) - \mu^2 = E(X^2) - E(X)^2$$

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

$V(CX) = E(C^2X^2) - E^2(CX) = C^2E(X^2) - [CE(X)]^2 = C^2V(X)$

V is not a linear operator.

Example: a random variable with an expectation value but without a variance.

$$P(x=n)=\frac{c}{n^3};;;E=\sum_n n\cdot\frac{c}{n^3}=c\sum_n\frac{1}{n^2}\leftarrow \text{converges}$$

$$E(X^2)=\sum_n n^2\cdot\frac{c}{n^3}=c\sum_n\frac{1}{n}\leftarrow \text{diverges}$$

Independence: random variables X and Y defined on the same space are called independent if

$$P(X,Y)=P(X)P(Y)$$

Tossing a coin twice: (1,1), (0,1), (0,0), (1,0) – the probability of the second toss is independent of the first.

In a product space: $P(X_1, X_2 \cdots, X_n)=P(X_1)P(X_2)\cdots\cdots P(X_n)$

$P(X_1, X_2 \cdots, X_n)$ – Joint probability

Uncorrelated random variables

If X and Y are independent random variables [$P(x_i y_j)=P(x_i)P(y_j)$] defined on the same sample space ⇕ they are uncorrelated, i.e.

$$E(X \cdot Y)=E(X) \cdot E(Y)$$

Proof: $E(X \cdot Y) = \sum_{ij} x_i y_j P(x_i y_j) = \sum_{ij} x_i y_j P(x_i)P(y_j) =$

$$= \sum_i x_i P(x_i) \cdot \sum_j y_j P(y_j) = E(X) \cdot E(Y)$$

X,Y independent ⇕ X,Y uncorrelated. The opposite is not always true. X,Y uncorrelated defined on the same sample space then

$$V(X+Y)=V(X)+V(Y)$$

Arithmetic average

$X_1, X_2, \ldots\ldots X_n$ are n random variables defined on the same sample space with the same expectation value $\mu = E(X_i)$, then the arithmetic average:

$$\bar{X} = \frac{X_1 + X_2 + \ldots . X_n}{n}$$

is also a random variable with $E(\bar{X}) = \mu$

$$E(\bar{X}) = E(\frac{X_1 + X_2 + \ldots X_n}{n}) = \frac{1}{n} E(X_1 + X_2 + \ldots X_n) =$$

$$= \frac{1}{n}[E(X_1) + E(X_2) \ldots + E(X_n)] = \frac{n\mu}{n} = \mu$$

Notice: $\bar{X}$ is defined over the product space $(x_1, x_2, \ldots\ldots x_n)$ with $P(x_1, x_2, \ldots\ldots x_n)$. $\bar{X}$ is important in simulations.

Variance of the arithmetic average

$X_1, X_2, \ldots\ldots, X_n$ are uncorrelated random variables with the same ❍ and $\blacklozenge^2$ ❶

$$\sigma^2(\overline{X}) = V(\overline{X}) = V\left(\frac{X_1 + X_2 + \ldots X_n}{n}\right) =$$

$$\frac{1}{n^2} V(X_1 + X_2 + \ldots X_n) = \frac{1}{n^2} \cdot nV = \frac{V}{n} = \frac{\sigma^2}{n}$$

❶ While the expectation value of the arithmetic average is also ❍, the variance decreases with increasing *n*!

The above result, $\sigma(\overline{X}) = \frac{\sigma}{\sqrt{n}}$ is extremely important, playing a central role in statistics & analysis of simulation data.

Sampling

So far we have dealt with probability spaces (ideal world), where the probability of an elementary event is known exactly and probabilities of events A could be calculated.

We defined the notion of a random variable (X) which is a function from the objects of a sample space to the real line, where the function can be defined according to the problem of interest. Cumulative distribution function and probability density function (for a continuous random variable) were defined. This enables one, at least in principle, to calculate expectation values $E(X)$ and variances $V(X)$ (and other parameters).

However, in most cases the probabilities in space are unknown and an analytical evaluation of the integrals $E(X)$ and $V(X)$ is unfeasible. Then one resorts to the "experimental" world on a computer – i.e., simulation where these parameters can be estimated by sampling techniques.

Examples where the probabilities are a-priori unknown

A. An uneven coin with unknown P and 1-P.

B. In statistical mechanics a real system is modeled by a probability space where the elementary events are the system configurations i with potential energy E_i and Boltzmann probability P_i^{B}

$$P_i^{\mathrm{B}} = \frac{\exp[-E_i / k_B T]}{Z} \qquad Z = \sum_i \exp[-E_i / k_B T]$$

P_i^{B} is unknown in most cases because Z is practically impossible to calculate.

C. Parameters of the normal distribution

The normal (Gaussian) distribution is very important. (p. 28) - defined by $E(\bigcirc)$ and V, which are not always known a-priori.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp -\left[\frac{x^2}{2\sigma^2}\right]$$

It can be shown that $f(x)$ is normalized,

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp -\left[\frac{x^2}{2\sigma^2}\right]dx = 1$$

and its expectation value is 0, because f is symmetric & x is an odd function

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = 0$$

The variance is therefore:

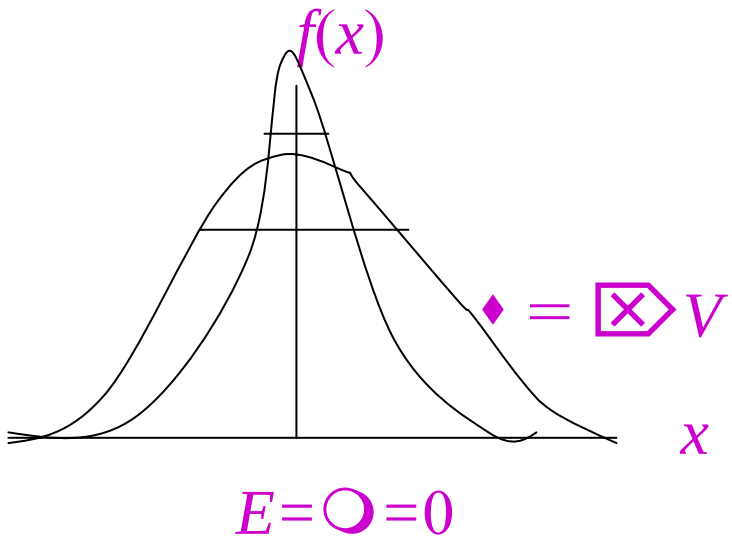
$$V(X) = \int_{-\infty}^{\infty} x^2 f(x)dx =$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 \exp-[\frac{x^2}{2\sigma^2}]dx = \frac{\sqrt{\pi}(2\sigma^2)^{\frac{3}{2}}}{2} \frac{1}{\sigma\sqrt{2\pi}} = \sigma^2$$

Here we used the integral:

$$\int_{-\infty}^{\infty} x^2 \exp-[ax^2]dx = \frac{\sqrt{\pi}}{2a^{3/2}}$$

Thus, a Gaussian is defined by *only* two parameters, $E(X)$ and $V(X)$ - in the above case, 0 and $\blacklozenge^2$, respectively. In the general case, $f(x) \sim \exp-[(x-\mathcal{O})^2/(2\blacklozenge^2)]$. $\mathcal{O}>0$ and $\blacklozenge$ defines the width of the distribution.



The integral of $f(x)$ from ❍ - ♦ ⓞ xⓞ ❍ + ♦ provides ~ 68% of the total probability = 1; the integral over ❍ - 2♦ ⓞ xⓞ ❍ + 2♦ covers ~95% of the total area.

Unlike the ideal case (i.e., known probabilities) a distribution might be known to be Gaussian but ❍ and ♦ are unknown. To estimate *them* one can sample x values from the distribution – the smaller is ♦ the larger the chance that x is closer to ❍.

Thus, while it is difficult to obtain the probability values, one can estimate $E(\mathcal{O})$ and V by sampling with the corresponding distribution without the need to know its values!

One example: a coin with unknown p (1) and 1-p (0); tossing this coin will produce a sample of relative frequencies ■(1) ✸p, ■(0) ✸1-p.

Another example: assume a container with hard spheres is shaken in space. While the spheres are distributed randomly we do not know the numerical value of this distribution.

Monte Carlo methods used on a computer enable one to sample with the Boltzmann prob. without knowing its value.

To estimate an unknown $E(X)$ in a rigorous and systematic way one uses the structure of the product space and the properties of the arithmetic average, $\overline{X} = \frac{X_1 + X_2 + \ldots X_n}{n}$. Thus, if $X_1, X_2, .., X_n$ are n equal uncorrelated random variables with ❍ and $\blacklozenge^2$ then $E(\overline{X}) = \mu$ and the variance decreases with n

$$V(\overline{X}) = \frac{\sigma^2}{n}$$

One can sample independently n times from this distribution generating one term (vector) $(x_1, x_2, \ldots\ldots, x_n)$ of the product space. There is a good chance that $(x_1 + x_2 + \ldots\ldots + x_n)/n$ will fall within one standard deviation, $V^{1/2}$ of the correct ❍. The larger is n the smaller is $V^{1/2}$ and the closer $(x_1 + x_2 + .. + x_n)/n$ should be to ❍. For n ✸ ◎ the estimation becomes exact since $V(\overline{X}) \to 0$

.

One example is a coin with unknown p (1) and 1-p (0), where ❍ $=E(X) = p$ and $\sigma^2=p(1-p)$ are unknown a-priori.

One can toss this coin n times independently. The result of this experiment is a single vector [e.g., (1,0,0,1….1,0)] out of the 2^n vectors of the product space. Estimation of ❍ is improved as n is increased due to the decrease of $V(\overline{X})$.

Also, $V(\overline{X})$ will decrease if σ^2 is decreased. If p ✹0 or p ✹1, $\sigma^2=p(1-p)$ ✹ 0 and even a single tossing experiment would lead (with high chance) to 0 and 1, respectively, and thus to the correct value of $E(X)$.

Thus, to estimate ❍ (and other properties) one has to move to the experimental world (sometimes on a computer) using the structure of probability spaces (p.9). Notice again that while the value of P [or $f(x)$] is unknown, one should be able to sample according to P! (see the above example for the coin).

This is the basic theory of sampling that is used in Monte Carlo and molecular dynamics simulations. However, notice that with these methods the random variables in most cases are correlated; therefore, to use the equation $V(\overline{X}) = \frac{\sigma^2}{n}$, the #

of samples generated, *n'* should be larger, sometimes significantly larger than *n*, the number of uncorrelated samples used in the above equation. This topic will be discussed in more detail later.

Monte Carlo simulations – the Metropolis method

Markov chains

We have discussed n independent experiments of tossing a coin, with elementary events (outcomes) i and j, i.e., the probabilities p_i and $p_j=1-p_i$ of experiment t do not depend on the outcome of experiment t-1. Assume a different situation where $p_i(t)$ depends on the outcome at t-1.

t-1	t	
i ⬇	$p(j\|i)$; $p(i\|i)$	$p(j\|i)+ p(i\|i)=1$
j ⬇	$p(j\|j)$; $p(i\|j)$	$p(j\|j)+ p(i\|j)=1$

The $p(i|j)$ are conditional probabilities and the process is called a Markov chain.

For simplicity we change the notation $p(i|j) = p(j \to i) = p_{ji}$, where p_{ji} is a transition probability - the probability to obtain i after having j in the previous experiment.

Assume now that only one experiment was carried out, such that at time 0 we have P_i^0 and P_j^0 where $P_i^0 + P_j^0 = 1$ - like for an independent coin. What is P_i^1 and P_j^1 of the next experiment?

$$P_i^1 = P_i^0 p_{ii} + P_j^0 p_{ji}$$

$$P_j^1 = P_i^0 p_{ij} + P_j^0 p_{jj}$$

or in a matrix form:

$$\left(P_i^1, P_j^1\right) = \left(P_i^0, P_j^0\right) \begin{pmatrix} p_{ii} & p_{ij} \\ p_{ji} & p_{jj} \end{pmatrix} \longleftarrow M$$

where M is a stochastic matrix – all its components ¤ 0 & the sum of probabilities in a line=1 ($p_{ii} + p_{ij} = 1; p_{ji} + p_{jj} = 1$)

One can ask: What is P_i^2 and P_j^2 - the same procedure:

$$\left(P_i^2, P_j^2\right) = \left(P_i^1, P_j^1\right) M = \left(P_i^0, P_j^0\right) M^2$$

and in general, after *n* experiments:

$$\left(P_i^n, P_j^n\right) = \left(P_i^0, P_j^0\right) M^n = \left(P_i^0, P_j^0\right) \begin{pmatrix} p_{ii} & p_{ij} \\ p_{ji} & p_{jj} \end{pmatrix}^n$$

This derivation is performed in the probability space (a product space of *n* experiments) with transition probabilities. The significance of this result in the experimental world is the following: If one repeats the series of *n* experiments many times always starting from P_i^0 and P_j^0, the number of times *i* and *j* will be obtained at the n^{th} experiment is proportional to - P_i^n and P_j^n, respectively.

One can define Markov chains with a larger number of states (EE) than 2 defined for a coin. If the number of states is N, the stochastic matrix will be of size $N \diamondsuit N$.

The interesting question is whether for a large number of experiments, i.e., a long Markov chain, the set of probabilities $\{P_i^n\}$ converge to values that are not changed for larger n and are independent of the initial set $\{P_i^0\}$.

The answer is positive. For "well behaved" Markov chains (i.e., all states are irreducible and a-periodic – not explained here), one obtains a unique set of stationary probabilities $\{\pi_i\}$

$$\lim_{n\to\infty} P_i^n = \pi_i \quad \text{(independent of the initial } \{P_i^0\}\text{.}$$

$$\pi_i > 0; \qquad \sum_i \pi_i = 1; \qquad \pi_j = \sum_i \pi_i p_{ij} \qquad [P_j^{n+1} = \sum_i P_i^n p_{ij}]$$

The existence of stationary transition probabilities is important. First, it is easy to carry out realizations of n experiments of a Markov chain using a computer (a realization is one term in the product space). One only has to know the matrix M of transition probabilities. A realization is obtained with the help of a random number generator.

The fact that $\{P_i^n\}$ converges to $\{\pi_i\}$ for a large n means that for $n' > n$ the states are chosen with $\{\pi_i\}$ - so, one can estimate expectation values defined with these stationary probabilities of random variables, such as the energy, from a single realization of the Markov chain (one term of the product space) in the same way it is done in the case of independent sampling (e.g., a regular coin or a die).

Two main points differentiate between independent and dependent sampling with a Markov chain. In an independent sampling of a coin, for example, all the experiments starting from the first one are considered, and due to independence the variance of any random variable decreases as $\blacklozenge_n = \blacklozenge/n^{1/2}$.

In the case of a Markov chain one has to ignore the first m experiments, where the probabilities relax to their stationary values. Also, because the experiments are not independent the random variables are correlated. If the correlation disappears after l experiments the variance decreases slower than in an independent sampling, i.e, only as $\blacklozenge_n = \blacklozenge/(n/l)^{1/2}$.

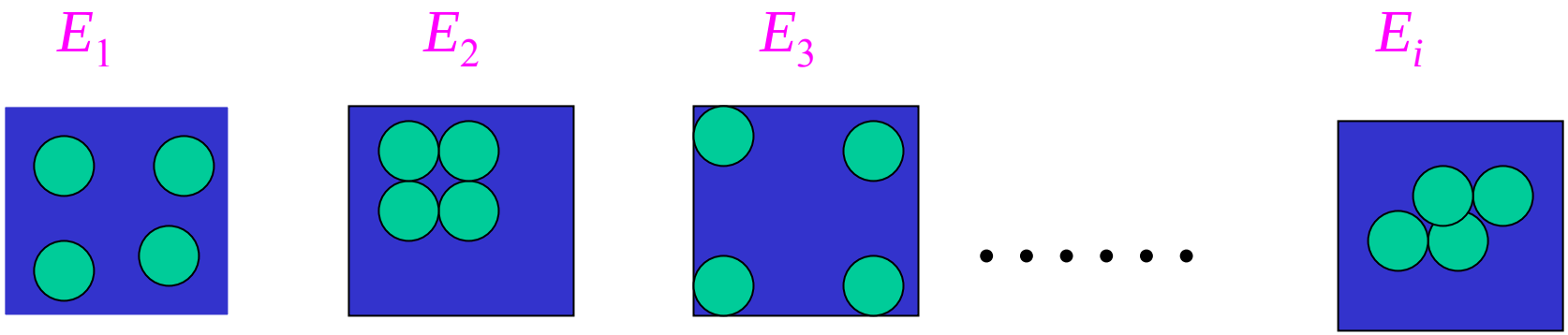
So, why to use a Markov chain if it is inferior to independent sampling? The reason is that a Markov chain enables one to sample with the Boltzmann probability $P_i^{\mathrm{B}} = \pi_i$.

Statistical mechanics

Treating a system by statistical mechanics requires defining a probability space based on the Boltzmann probability, P_i^{B}, which in most cases cannot be calculated.

However, using a Markov chain, system configurations can be sampled according to P_i^{B} without knowing its values (the Metropolis Monte Carlo method).

The (*NVT*) ensemble – probability space - *all* possible configurations:



E_i – potential energy of config i (random variable); T- absolute temp, k_B - Boltzmann constant. The Boltzmann probability and partition function Z:

$$P_i^{\mathrm{B}} = \frac{\exp[-E_i / k_B T]}{Z} \qquad Z = \sum_i \exp[-E_i / k_B T]$$

P_i^{B} is unknown because Z is very difficult to calculate. Most of the contribution to Z and P_i^{B} come from a very small region. We are interested, for example, in the average potential energy,

$$\langle E \rangle = \sum_i P_i^B E_i$$

With a Markov chain one can find the extremely small region in configurational space that provides most of the contribution to the partition function, & sample it.
This is the Metropolis Monte Carlo (MC) method described below first as applied to an 1*d* Ising model. Before discussing MC we describe an important tool for simulation – the random number generator.

Random number generators

Suppose a probability space consisting of two elementary events *i* and *j* with probabilities 0.2 and 0.8, respectively. We want to carry out *n* independent experiments (product space) using a computer. We use a program called random number generator that provides numbers within (0,1] distributed uniformly.

For each experiment a new random number r is generated:
If $r \leq 0.2$ event i is chosen; if $r > 0.2$, j is chosen.

Main Program

```
real*8 seed
seed=8957321.d0
.
x=random(seed)          random(seed) is 0 < x ≤1
.
end
```

Function random(seed)

```
seed=mod(69069.d0*|seed|+1.d0,2^32)
random=seed/2^32                    2^32= 4,294,967,296.d0
return; end
```

mod(a,b) is the remainder of a/b, e.g., mod(17,5)=2

The Ising model defines a probability space. The model is defined on a linear lattice of N sites. At each lattice site k there is a spin that can be in two states, $\sigma_k = +1$ or -1.

$+ - - - + - + + - - - -$ ……

So, there are 2^N possible spin configurations (elementary events). Two *nearest neighbor* spins, k and l interact with energy, $\mathcal{H}_{kl} = -J\sigma_k \sigma_l$ $(J>0)$ $\mathcal{H}_{++} = \mathcal{H}_{--} = -J$; $\mathcal{H}_{+-} = \mathcal{H}_{-+} = +J$.

The total energy of spin configuration i, $E_i = \Sigma_{kl} \mathcal{H}_{kl}(i)$ is a random variable. The (Boltzmann) probability of i is,

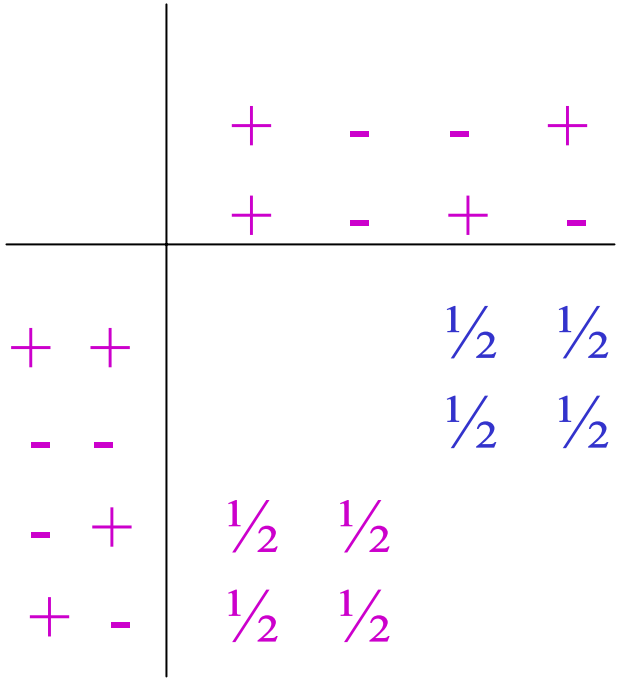
$$P_i^{\mathrm{B}} = \exp[-E_i/k_{\mathrm{B}}T]/Z \qquad Z = \Sigma_i \exp[-E_i/k_{\mathrm{B}}T]$$

k_{B} – Boltzmann constant; T - absolute temperature, Z- partition function (this model can be solved analytically).

We return now to the MC method applied for simplicity to the 1d Ising model at temperature T. We seek to generate a sample of configurations distributed with the Boltzmann probability.

Our mission is to define a Markov chain that its stationary probabilities are Boltzmannian, $\pi_i = P_i^B = \mathbf{exp}[-E_i / k_B T] / Z$. Ising has 2^N states (as compared to 2 states of a coin). We define transitions only between states that differ by a single spin. For example: + - + ✸ + + + therefore instead of the 2^N possible transitions only N are allowed.

Staying in state i we define the same trial transition probability (TP) $T_{ij} = 1/N$ to go to any of the N allowed states j. For two spins the number of states is of $2^2=4$ and a symmetric matrix $T_{ij} = 1/2$ is obtained.



$T_{ij}=T_{ji}$

For every state transition is defined into two states with TP=1/2. The process starts from an arbitrary configuration *i*. A site *k* is selected with $T_{ij}=1/N$ - Reversing the sign of the spin on the selected site defines configuration *j*. Second stage: a decision should be made whether to accept *j*, i.e., to flip the spin on site *k*. This is done with p_{ij} which was suggested by Metropolis, Rosenbluth[2] and Teller[2] (1953).

They suggested the following prescription which is also called the detailed balance condition

$$\frac{T_{ij}\,p_{ij}}{T_{ji}\,p_{ji}} = \frac{p_{ij}}{p_{ji}} = \frac{P_j^{\mathrm{B}}}{P_i^{\mathrm{B}}} = \exp[-(E_j - E_i)/k_B T]$$

Notice that because of symmetry the ratio of the T_{ij}'s is 1 and there is a freedom to determine the p_{ij} because they are defined up to their ratio.

Indeed, P_i^{B} are stationary. Using detailed balance one obtains

$$\sum_j P_j^{\mathrm{B}} p_{ji} = \sum_j P_j^{\mathrm{B}} \cdot p_{ij} \frac{P_i^{\mathrm{B}}}{P_j^{\mathrm{B}}} = P_i^{\mathrm{B}} \sum_j p_{ij} = P_i^{\mathrm{B}} \cdot 1 = P_i^{\mathrm{B}}$$

Detailed balance is a sufficient condition for the Boltzmann probabilities to become stationary. Also, one has to verify that

T_{ij} is symmetric (very important) and the Markov chain is irreducible and aperiodic (which guarantees ergodicity). A popular set of p_{ij} is the following:

$$p_{ij} = \begin{cases} 1 & \Delta E_{ji} < 0 \\ \exp[-\Delta E_{ji} / k_B T] & \Delta E_{ji} \geq 0 \end{cases} \qquad \Delta E_{ji} = E_j - E_i$$

It is easy to see that $p_{ij}/\ p_{ji} = \exp\text{-}[(E_j\text{-}E_i)/k_B T]$, i.e., the detailed balance condition is satisfied.

$$\begin{matrix} E_1 & E_2 & E_3 \ldots\ldots & E_n \end{matrix}$$

$$\begin{pmatrix} 1 \\ 2 \\ 3 \\ \bullet \\ \bullet \\ \bullet \\ 2^N \end{pmatrix} \times \begin{pmatrix} 1 \\ 2 \\ 3 \\ \bullet \\ \bullet \\ \bullet \\ 2^N \end{pmatrix} \times \begin{pmatrix} 1 \\ 2 \\ 3 \\ \bullet \\ \bullet \\ \bullet \\ 2^N \end{pmatrix} \bullet\bullet\bullet \times \begin{pmatrix} 1 \\ 2 \\ 3 \\ \bullet \\ \bullet \\ \bullet \\ 2^N \end{pmatrix} \quad n \text{ times}$$

Performing the process is simple. One starts with an arbitrary configuration; using a random number, a site k is selected with $T_{ij}=1/N$. The energy $E(\blacklozenge_k)$ of spin $\blacklozenge_k$ with its nearest neighbor spins is calculated together with $E(-\blacklozenge_k)$; if $E(-\blacklozenge_k)- E(\blacklozenge_k)<0$ the spin is flipped to $-\blacklozenge_k$; in the other case a random number r is generated: if r ● $\exp-[(E_j-E_i)/k_BT]$, $\blacklozenge_k$ ✹ $-\blacklozenge_k$ otherwise $\blacklozenge_k$ remains and the process continues.

Comments

(1) A very simple method – minimal configurational changes.

(2) Configurations are selected with the Boltzmann probability ❶ properties such as the energy, that are directly measured on the system can be estimated by the arithmetic average calculated over a sample of size n, $1/n \blacklozenge_t E_{i(t)}$.

However, the value of P_i^{B} is unknown and it is difficult to obtain the absolute entropy in a direct manner because it requires, $1/n \blacklozenge_t \ln P_{i(t)}^{\mathrm{B}}$.

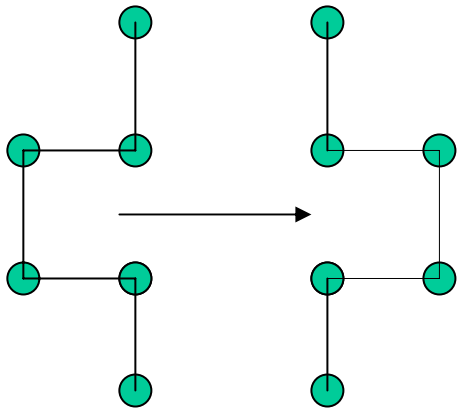
(3) The configurations are highly correlated which might need generating large samples.

(4) The MC method is a dynamical procedure that can model relaxation processes, for example.

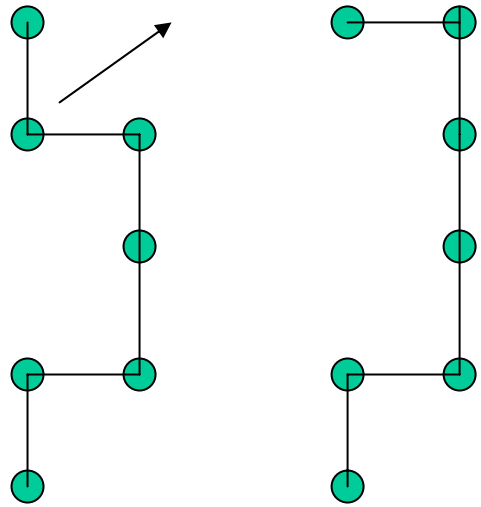
(5) The above MC is very efficient for liquids, magnetic systems, and some MC versions also for polymers.

It fails close to a phase transition, dense systems, and compact proteins (see later).

Application of MC to other systems requires defining the basic configurational change & verifying that the detailed balance condition is satisfied. For a self-avoiding walk (SAW) on a square lattice (no energy) one can define a crankshaft move and a single monomer flip.

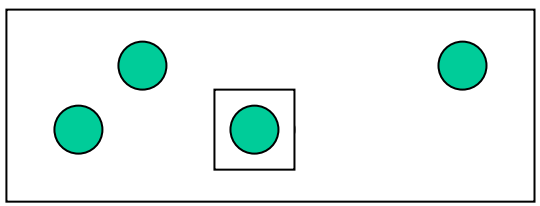


crankshaft



single monomer flip

For a Lennard-Jones liquid one defines a small box around a molecule. A trial move is chosen at random within this box.



If the system is dense and the box around the molecule is relatively large the trial move in most cases will not be accepted because there is a high chance that the particle will "bump" another one leading to high energy, E_j ✸ ◎ ✸ exp-$[(E_j-E_i)/k_BT]$ ✸ 0 ✸ no acceptance. Therefore, in this case the size of the box should be decreased until the acceptance rate is 30-40%.

However, very small moves will require a long simulation time to "scramble" the system, i.e., to obtain a reliable sample. Therefore, if the system is very dense almost nothing will happen ✸ MC becomes inefficient.

Thus, even though there is a mathematical guarantee that the simulation will converge to the stationary probabilities (P_i^B)

The relaxation time can be very large and the correlations large as well. No theorem (criterion) exists which would tell us when convergence has been reached.

Therefore, one looks for an apparent stability in the averages. A practical measure is to carry out several such runs, starting from different configurations and using different random number sequences - to verify convergence.

Notice, there are many other versions of the Metropolis MC, where larger clusters are changed, different T_{ij} and p_{ij} are defined, etc.

Problems:

1) A die is rolled four times. What is the probability to obtain 6 exactly one time? (answer: 0.3858)

2) A box contains 6 red balls, 4 white balls, and 5 blue balls. 3 balls are drawn successively. Find the probability that they are drawn in the order red, white, and blue if the ball is (a) replaced, (b) not replaced. (a: 0.0356; b:0.044)

3) What is the expectation value of m in the random variable of Poisson:

$$P(X = m) = \lambda^m \exp(-\lambda)/m! \quad (m = 0,1,2,\ldots..).$$

4) Show that the number of ways n objects can be divided into k groups of $r_1, r_2, \ldots..r_k$; $\swarrow\ r_k = n$ without considering the order in each group but considering the order between the groups is

$$n!/(r_1!r_2!\ldots.r_k!)$$ (see p. 14)

5) Two random variables X and Y are uncorrelated if $E(XY)=E(X)E(Y)$. Show that in this case:

$$V(X+Y)=V(X)+V(Y).$$

V is the variance.

6) Six passengers sit on a bus which visits 9 bus stops. What is the probability that two of the passengers will never get off the bus in the same bus stop? (0.14)

7) A die is rolled 36 times. What is the probability to obtain six times every number?

9) A person bought 6 tickets in a lottery of 30 tickets, which has 6 prizes. What is the probability that he wins at least one prize? (calculate the complementary event).0.227 (0.773)