Probability fundamentals for statistical mechanics

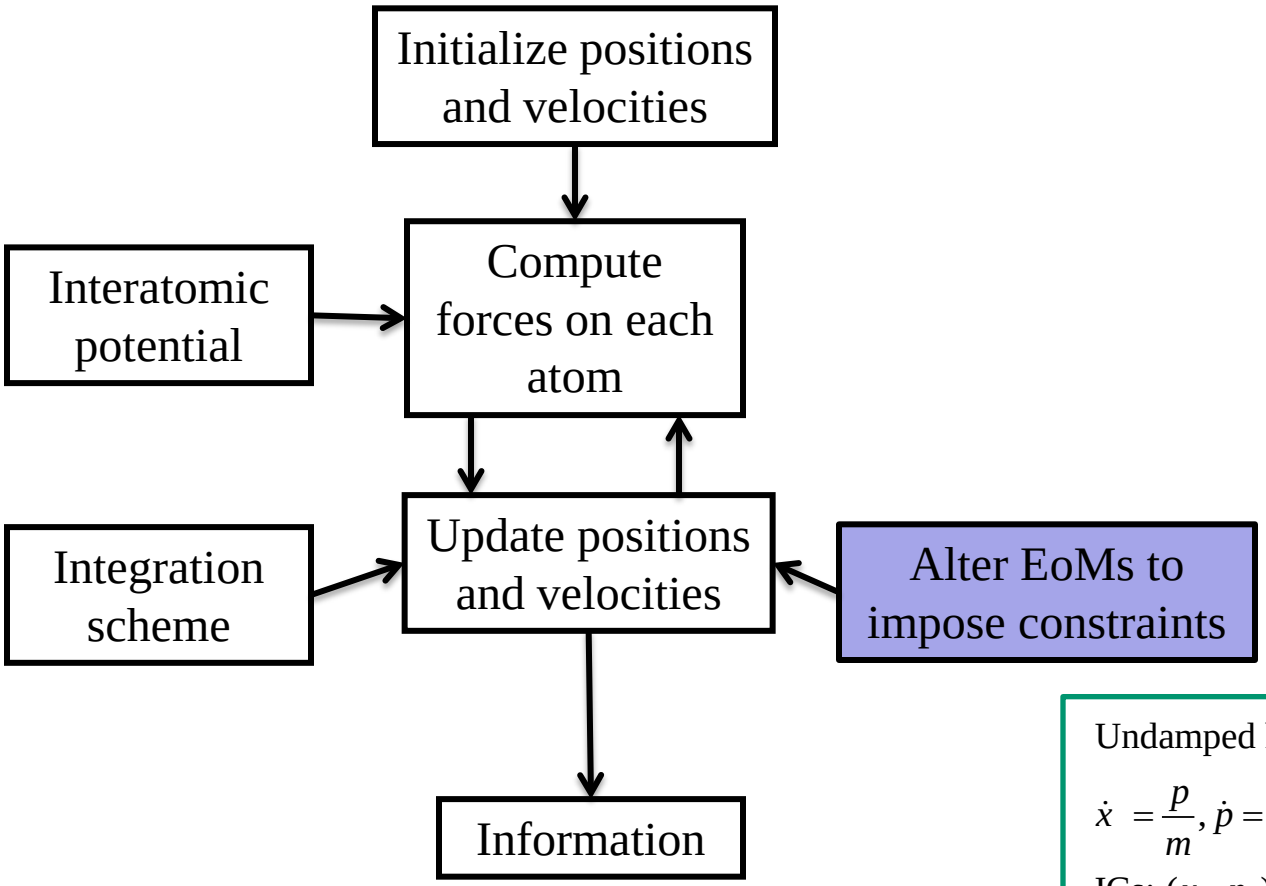
Baidurya Bhattacharya

GIAN Course on Nonequilibrium Molecular Dynamics and Statistical Mechanics

IIT Kharagpur, December 2016

Motivation

Molecular Dynamics



Hamiltonian Formulation:

$$\dot{x} = \frac{\partial H}{\partial p}, \dot{p} = -\frac{\partial H}{\partial x}$$

Generalize to $6N$ dofs:

$$\dot{x}_i = \frac{\partial H}{\partial p_i}, \dot{p}_i = -\frac{\partial H}{\partial x_i}$$

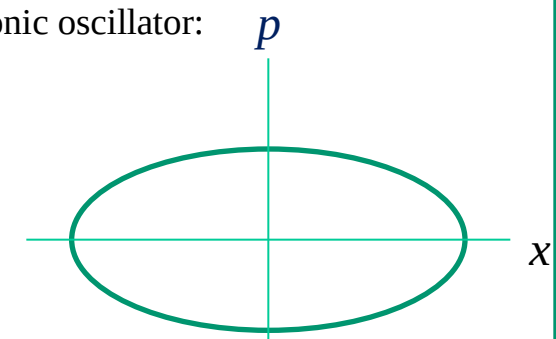
$$H = E = k\frac{x^2}{2} + \frac{p^2}{2m}$$

Undamped harmonic oscillator:

$$\dot{x} = \frac{p}{m}, \dot{p} = -kx$$

ICs: (x_0, p_0)

Phase space plot:



Conservative system

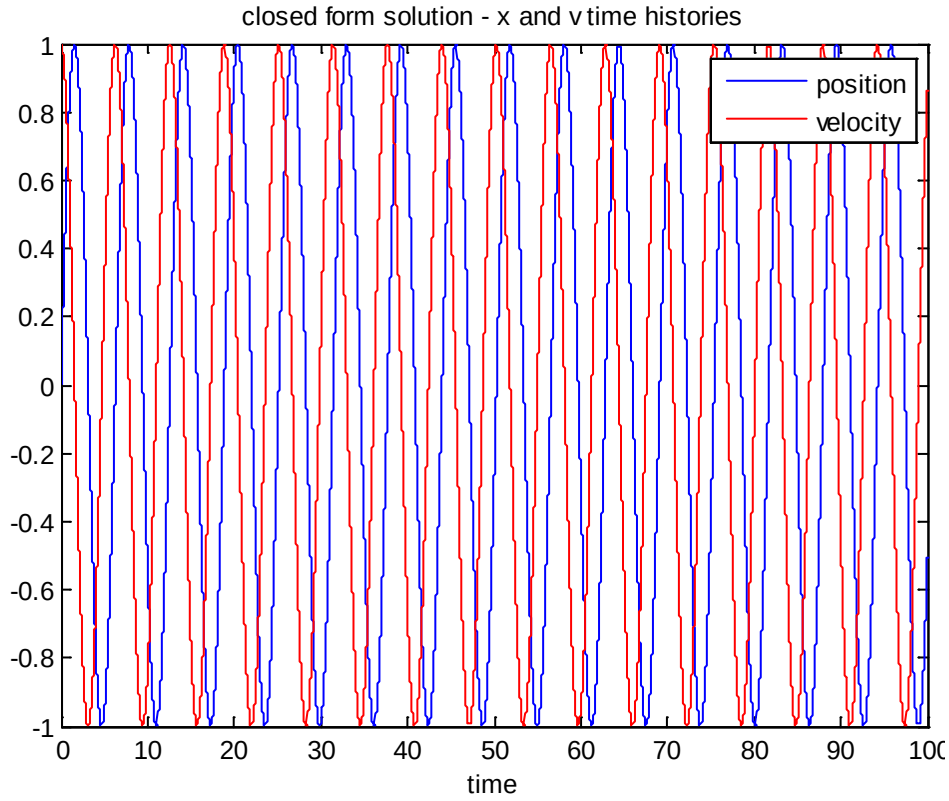
```
k=1;m=1;
w=sqrt(k/m);
%Define the initial conditions for the single
harmonic oscillator
y(1) = 0.0; %location
y(2) = 1.0; %velocity
E0=0.5*m*y(2)*y(2);
A=sqrt(2*E0/k);

dt = 0.001; %the step size
steps = 10000; %Total number of time steps

t=0;
for i=1:steps
    t=t+dt;
    x=A*sin(w*t);
    v=A*w*cos(w*t);
    cenergy(i)=0.5*k*x^2 + 0.5*m*v^2;
    ctime(i)=i*dt; %c=closed form
    cy1(i)=x;
    cy2(i)=v;

end
```

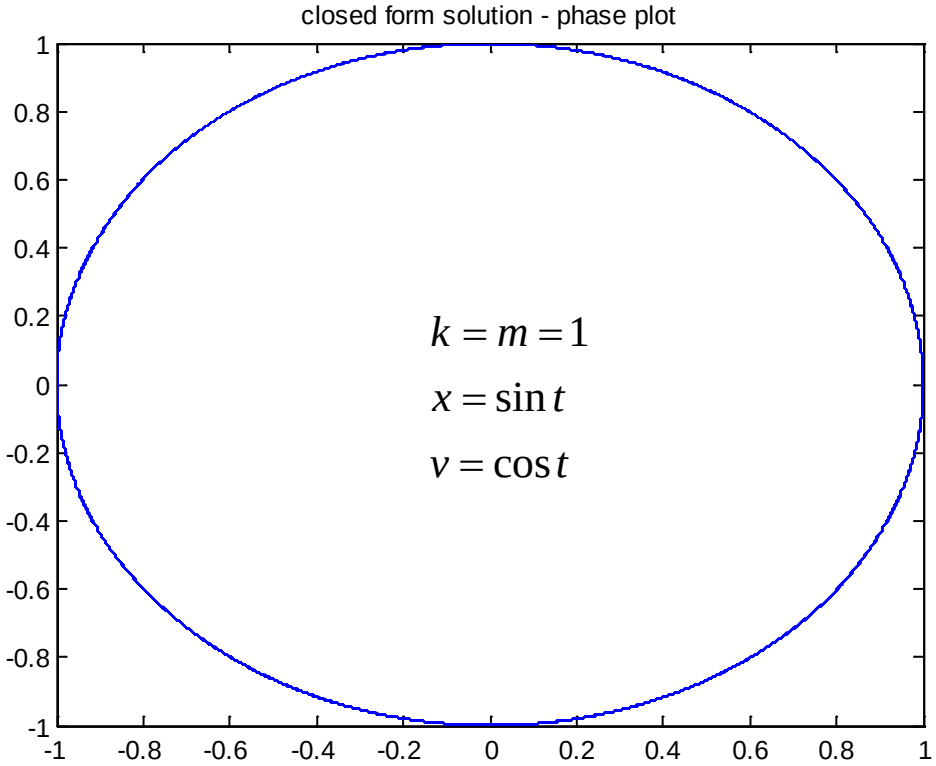
Undamped harmonic oscillator:

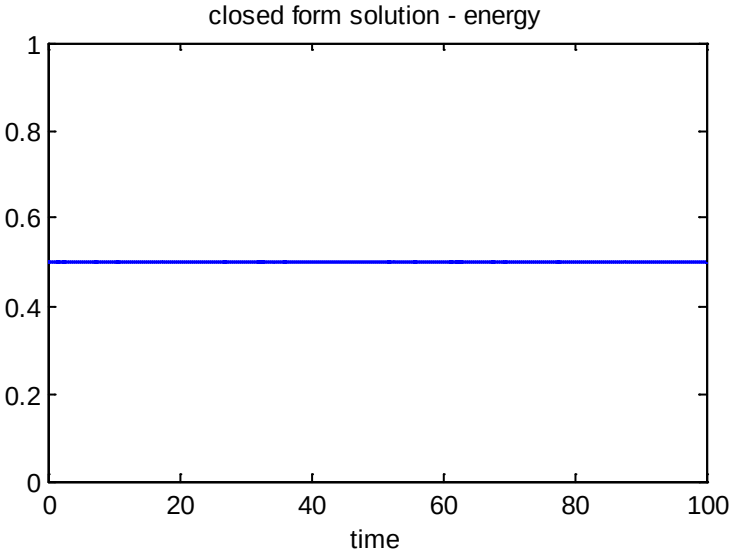
$$\dot{x} = \frac{p}{m}, \dot{p} = -kx$$

ICs: (x_0, p_0)

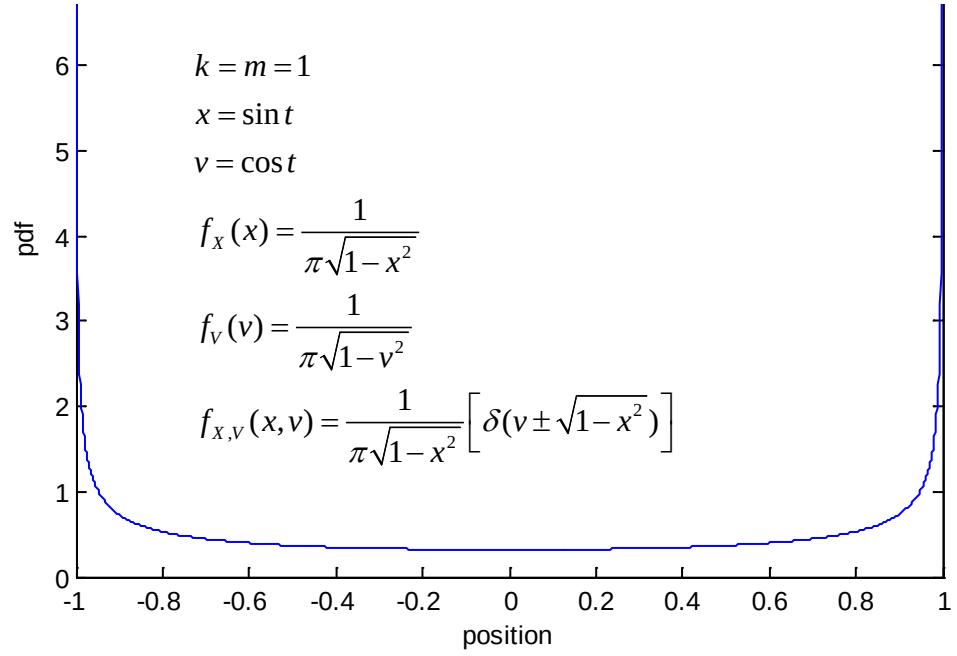
Microcanonical dynamics

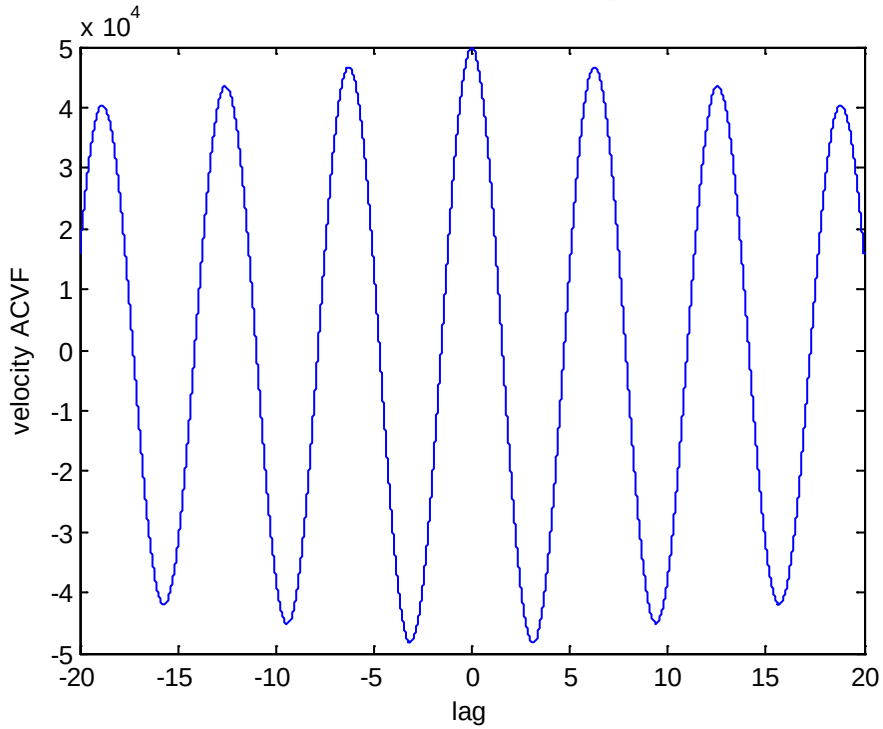


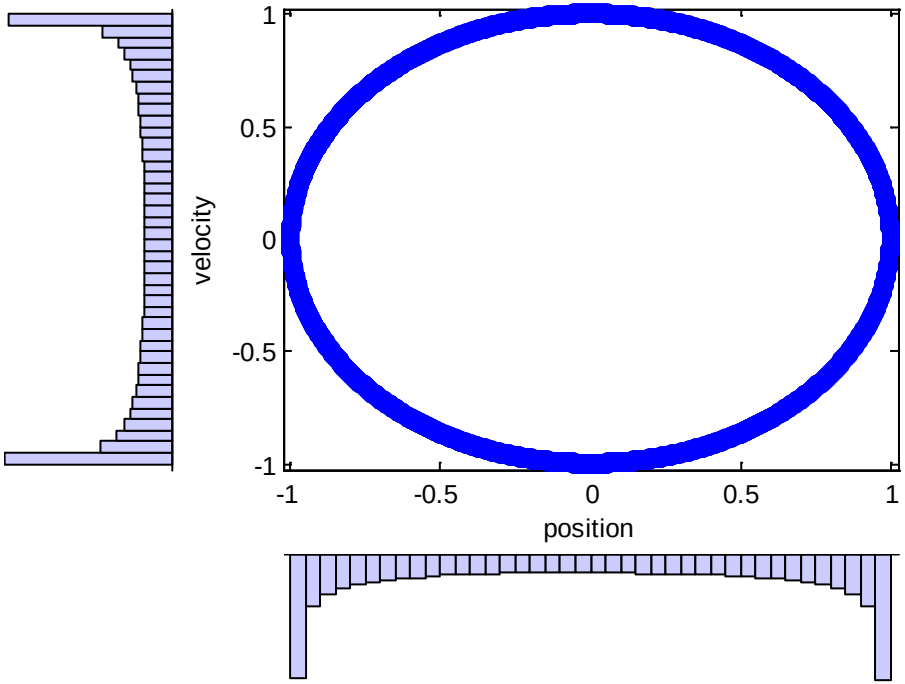


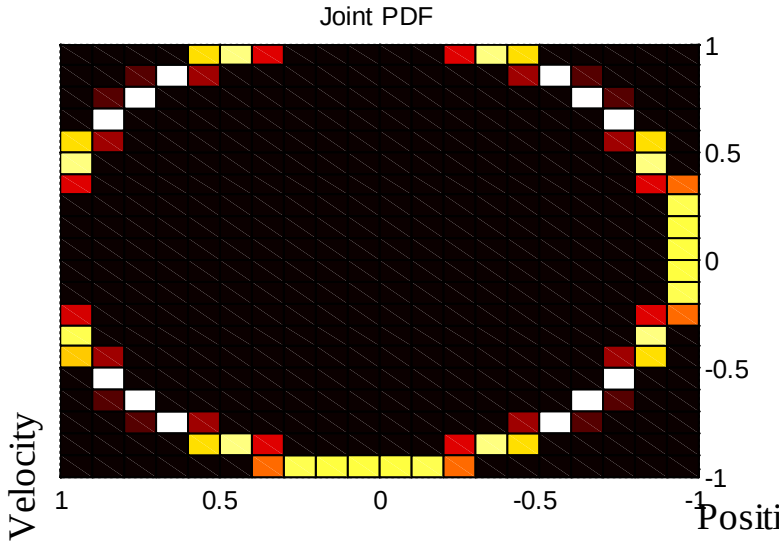


Microcanonical dynamics









Introducing dissipation

```
function rungekutta4_viscous
for(i=1:steps)
    t = t+dt;
    time(i)=t;
    y = Runge4(t,y,dt,N);
    energy(i) = 0.5*(k*y(1)^2+m*y(2)^2);
end

function y = Runge4(time,y,dt,N)
step = dt/2.0;

for(i=1:N)
    k1(i) = step*eom(time,y,i);
    t1(i) = y(i) + 0.5*k1(i);
end
for(i=1:N)
    k2(i) = step*eom(time+dt,t1,i);
    t2(i) = y(i) + 0.5*k2(i);
end
for(i=1:N)
    k3(i) = step*eom(time+dt,t2,i);
    t3(i) = y(i) + k3(i);
end
for(i=1:N)
    k4(i) = step*eom(time+dt,t3,i);
end
for(i=1:N)
    y(i) = y(i) +
(k1(i)+2.0*k2(i)+2.0*k3(i)+k4(i))/6.0;
end
end
```

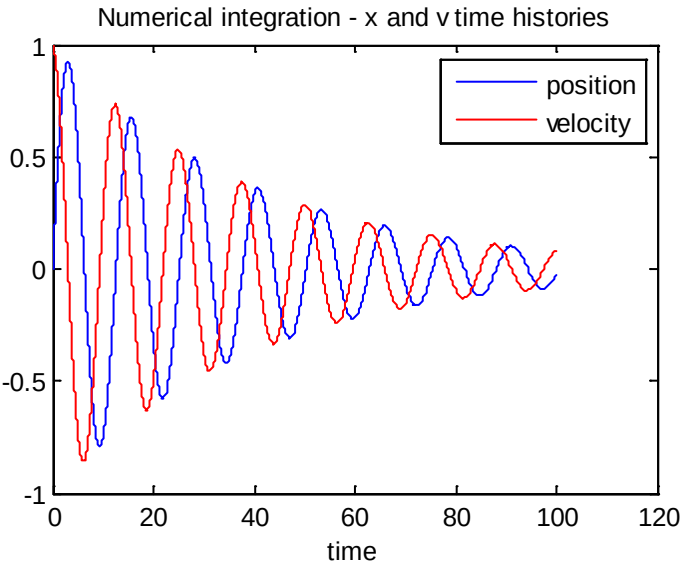
Damped harmonic oscillator:

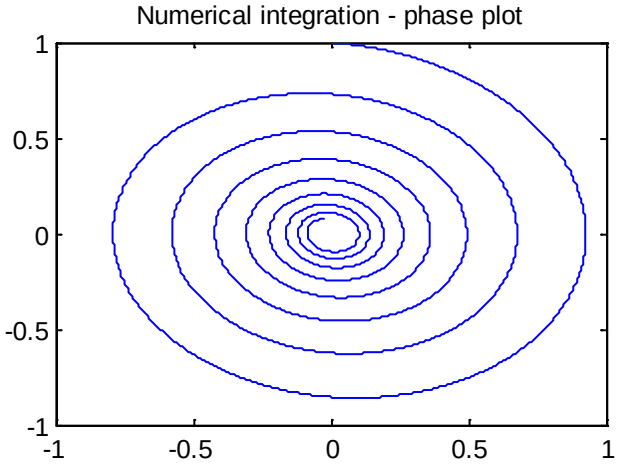
$$\dot{x} = \frac{p}{m}, \dot{p} = -kx - \eta p$$

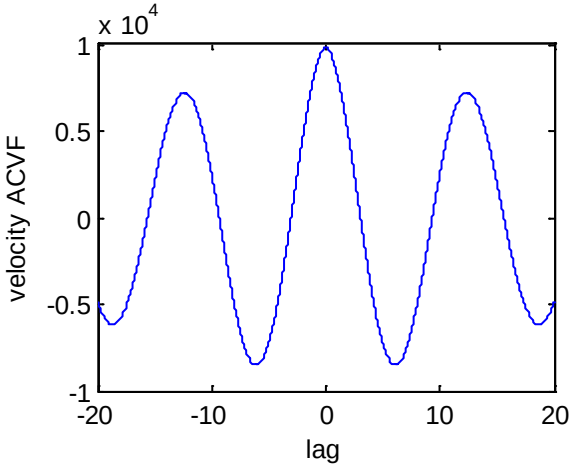
ICs: (x_0, p_0)

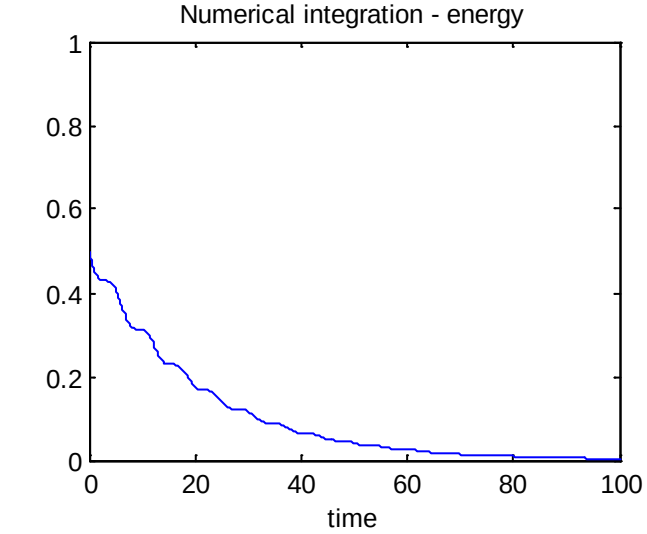
```
function dydt = eom(t,y,i)
if(i==1)
    dydt = y(2);
end
if(i==2)
    dydt = -k/m*y(1)-eta*y(2);
end
```

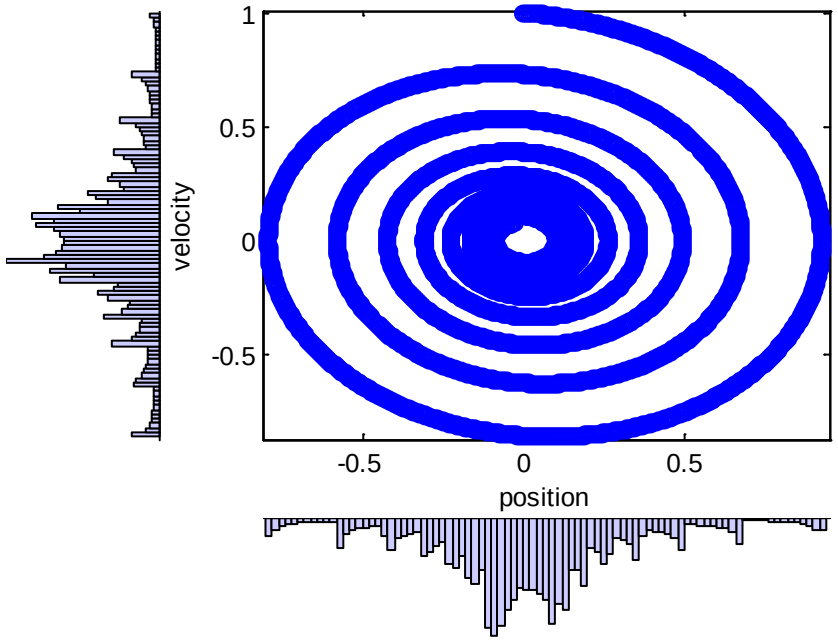
Viscous drag











Nose – Hoover thermostat

- Modified EoM:

$$\dot{x}_i = p_i$$

$$\dot{p}_i = -\nabla_i \Phi - \eta p_i$$

$$\dot{\eta} = Q_{NH}^{-1}\left[\sum_i p_i^2 - N k_B T\right]$$

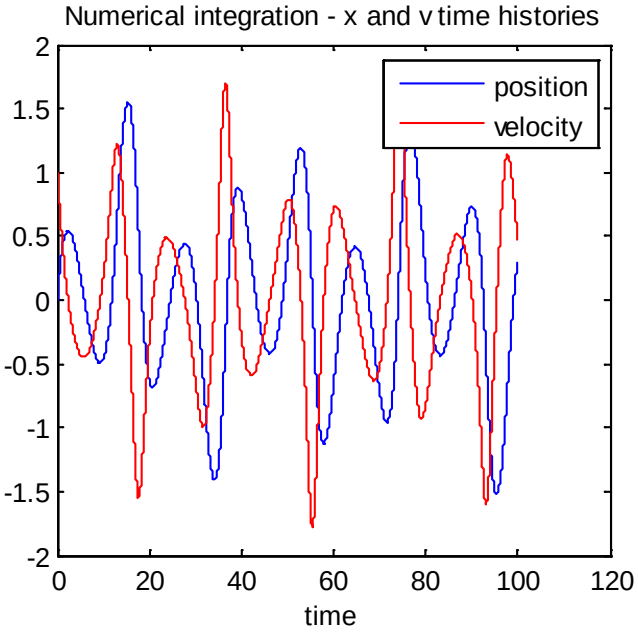
- Reservoir represented by the new variable
- Reservoir acts as "damping/exciting" force, withdrawing energy at times and supplying energy at others

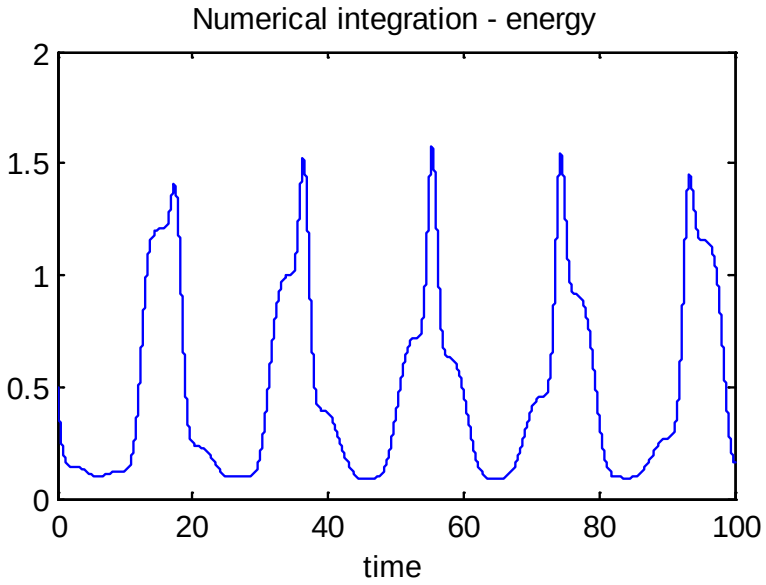
NH Dynamics

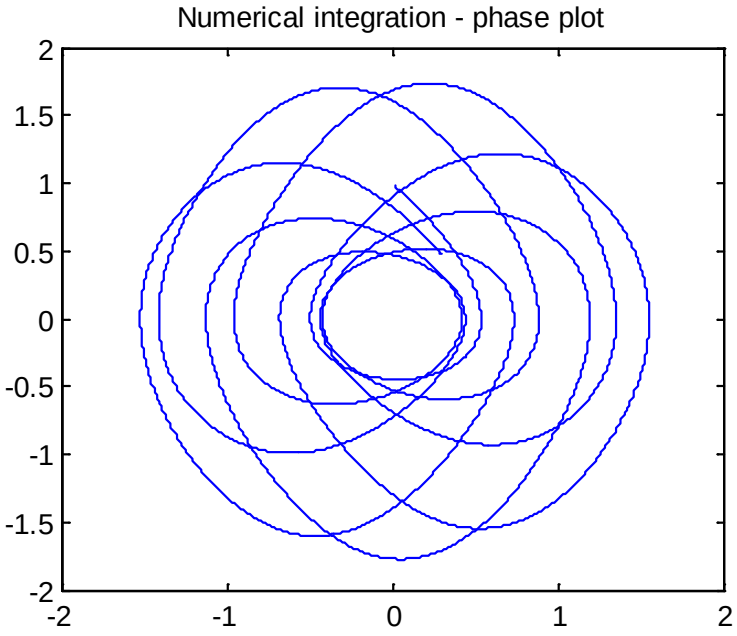
```
N=3; % no. of equations
k=1;m=1; Qeta=1;
dt = 0.001; %Defines the step size
steps = 100000; %Total number of time steps
y(1) = 0.0; %initial location
y(2) = 1.0; %initial velocity
y(3)=1; %initial eta

t = 0.0; %initializing time
for(i=1:steps)
    t = t+dt;
    time(i)=t;
    y = Runge4(t,y,dt,N);
    energy(i) = 0.5*(k*y(1)^2+m*y(2)^2);
    x1(i) = y(1);
    v1(i) = y(2);
end
```

```
function y = Runge4(time,y,dt,N)
step = dt/2.0;

for(i=1:N)
    k1(i) = step*eom(time,y,i);
    t1(i) = y(i) + 0.5*k1(i);
end
for(i=1:N)
    k2(i) = step*eom(time+dt,t1,i);
    t2(i) = y(i) + 0.5*k2(i);
end
for(i=1:N)
    k3(i) = step*eom(time+dt,t2,i);
    t3(i) = y(i) + k3(i);
end
for(i=1:N)
    k4(i) = step*eom(time+dt,t3,i);
end
for(i=1:N)
    y(i) = y(i) +
(k1(i)+2.0*k2(i)+2.0*k3(i)+k4(i))/6.0;
end
end

function dydt = eom(t,y,i)
if(i==1)
    dydt = y(2);
end
if(i==2)
    dydt = -k/m*y(1)-y(3)*y(2);
end

if(i==3)
    dydt=(y(2)^2-.5)/Qeta;
end
end
```

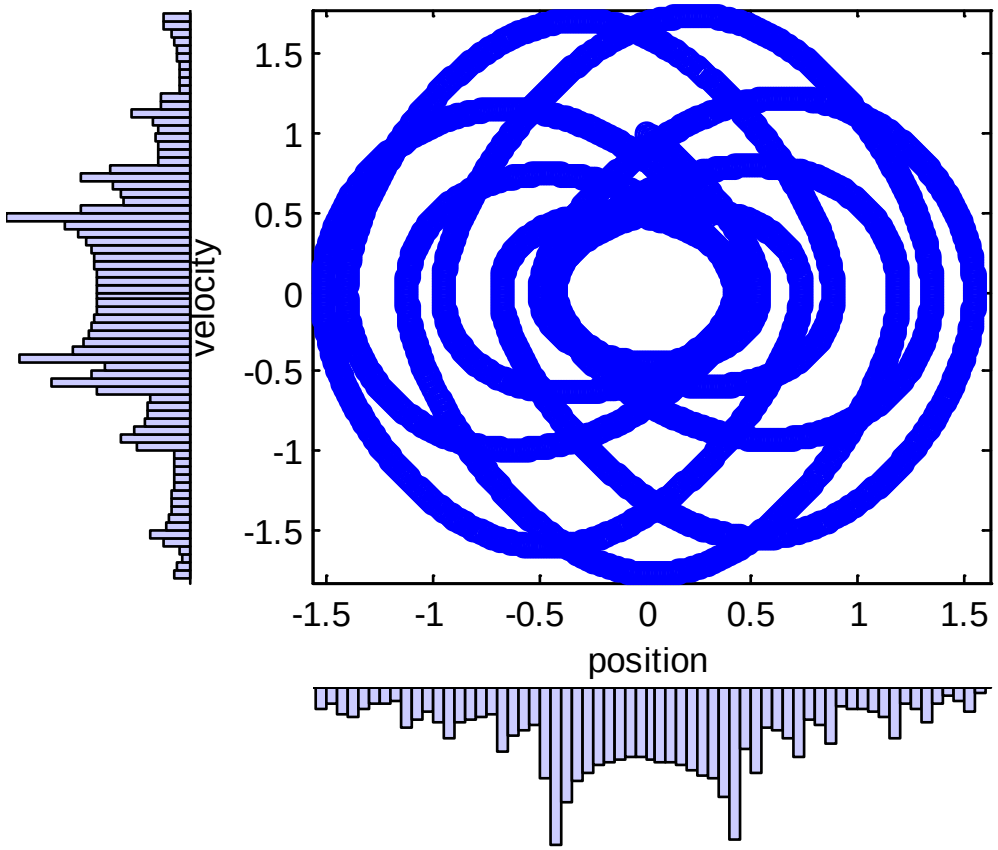
NH Dynamics

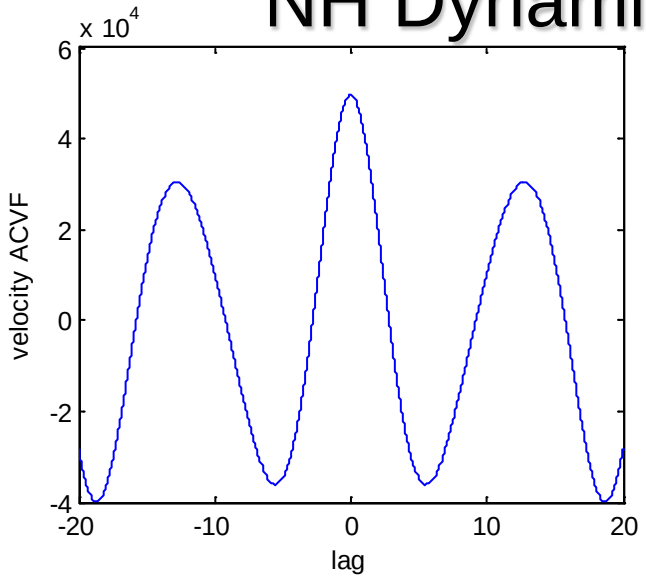

Numerical integration - x and v time histories
position
velocity
time

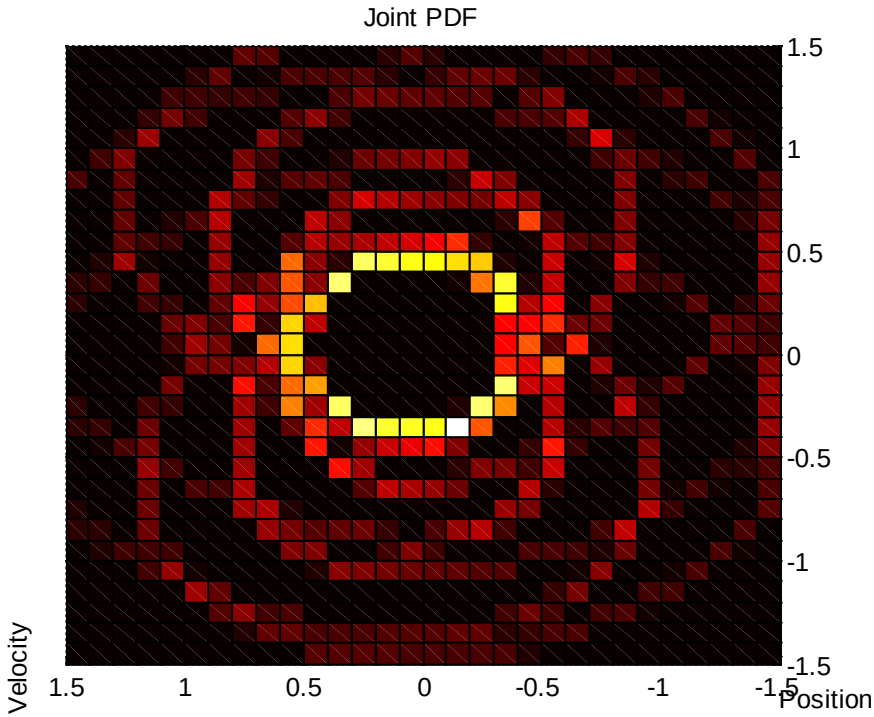

Numerical integration - energy
time


Numerical integration - phase plot

NH Dynamics


velocity
1.5
1
0.5
0
-0.5
-1
-1.5
-1.5
-1
-0.5
0
0.5
1
1.5
position


x 10^4
6
4
2
0
-2
-4
velocity ACVF
-20
-10
0
10
20
lag


Joint PDF
1.5
1
0.5
0
-0.5
-1
-1.5
Velocity
1.5
1
0.5
0
-0.5
-1
-1.5
Position

Langevin dynamics

```
for(i=1:steps)
    t = t+dt;
    time(i)=t;
    y = Runge4(t,y,dt,N);
    energy(i) = 0.5*(k*y(1)^2+m*y(2)^2);
end
```

$$\dot{x}_i = p_i / m$$

$$\dot{p}_i = -\frac{\partial \Phi}{\partial x_i} - \gamma p_i - \alpha \xi(t)$$

$$< \xi(t) >= 0$$

$$< \xi(t)\xi(t') >= \alpha \delta(t - t')$$

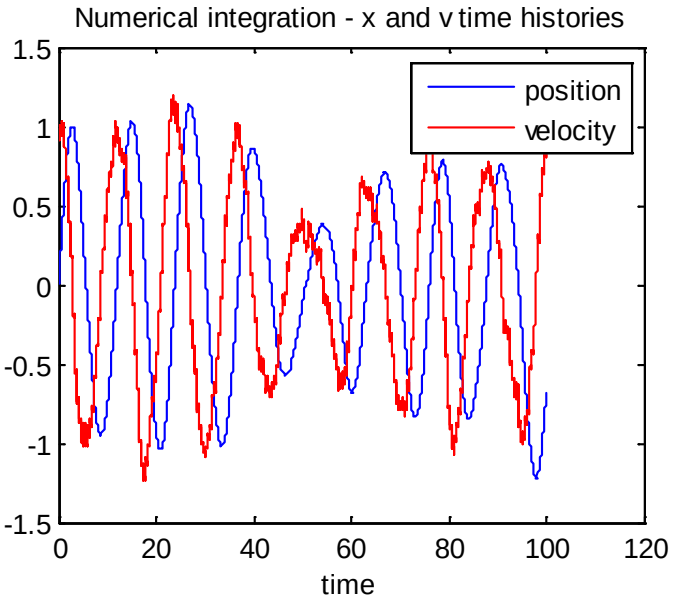
At equilibrium, using MB distribution:

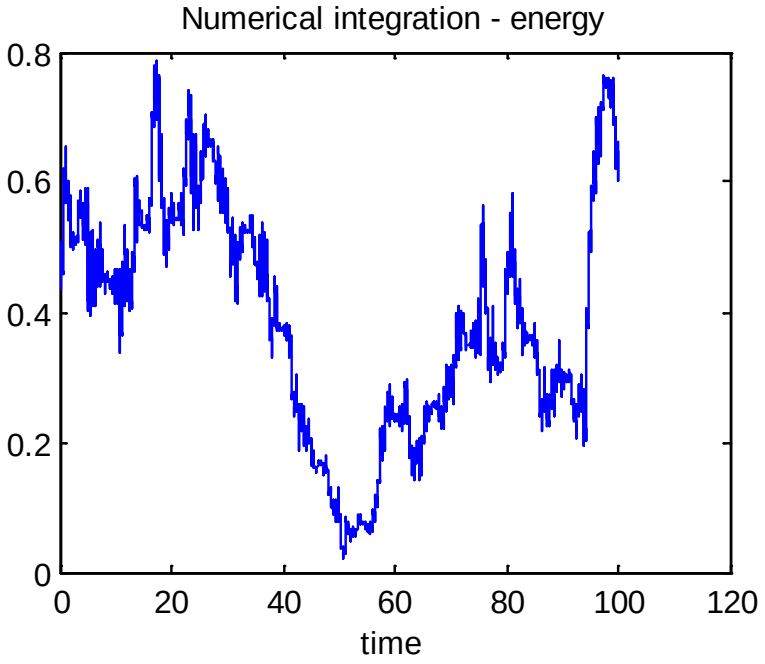
$$\alpha = \sqrt{2m\gamma k_B T}$$

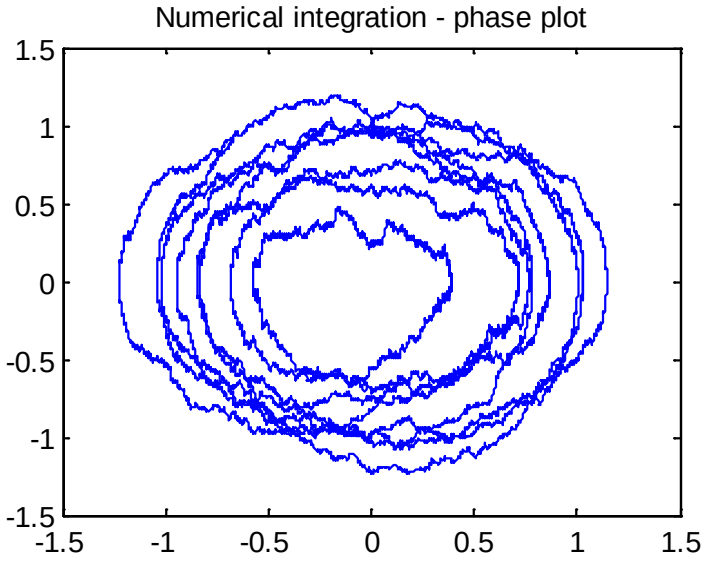
```
function y = Runge4(time,y,dt,N)
step = dt/2.0;

for(i=1:N)
    k1(i) = step*eom(time,y,i);
    t1(i) = y(i) + 0.5*k1(i);
end
for(i=1:N)
    k2(i) = step*eom(time+dt,t1,i);
    t2(i) = y(i) + 0.5*k2(i);
end
for(i=1:N)
    k3(i) = step*eom(time+dt,t2,i);
    t3(i) = y(i) + k3(i);
end
for(i=1:N)
    k4(i) = step*eom(time+dt,t3,i);
end
for(i=1:N)
    y(i) = y(i) +
(k1(i)+2.0*k2(i)+2.0*k3(i)+k4(i))/6.0;
end
end
```

```
function dydt = eom(t,y,i)
global eta sqrt2etakT dt k m Qeta xmax
if(i==1)
    dydt = y(2);
end
if(i==2)
    dydt = -k/m*y(1)-
eta*y(2)+sqrt2etakT*randn/sqrt(dt/2);
end
end
```

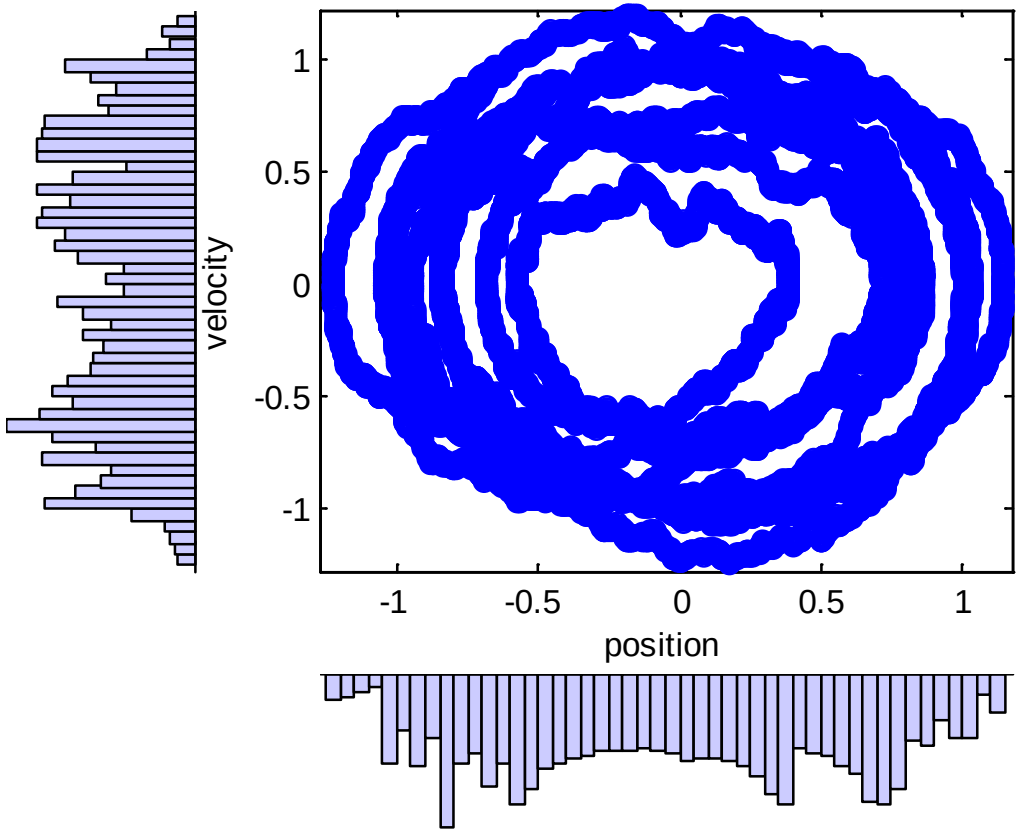
Langevin dynamics

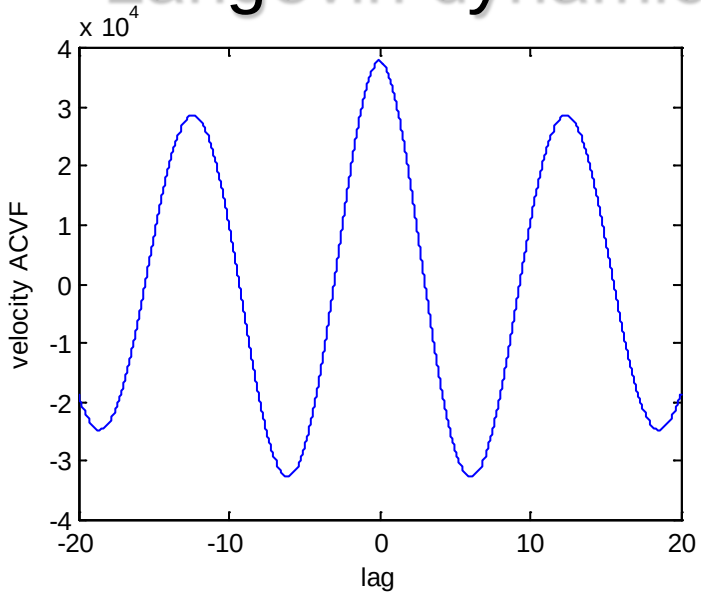


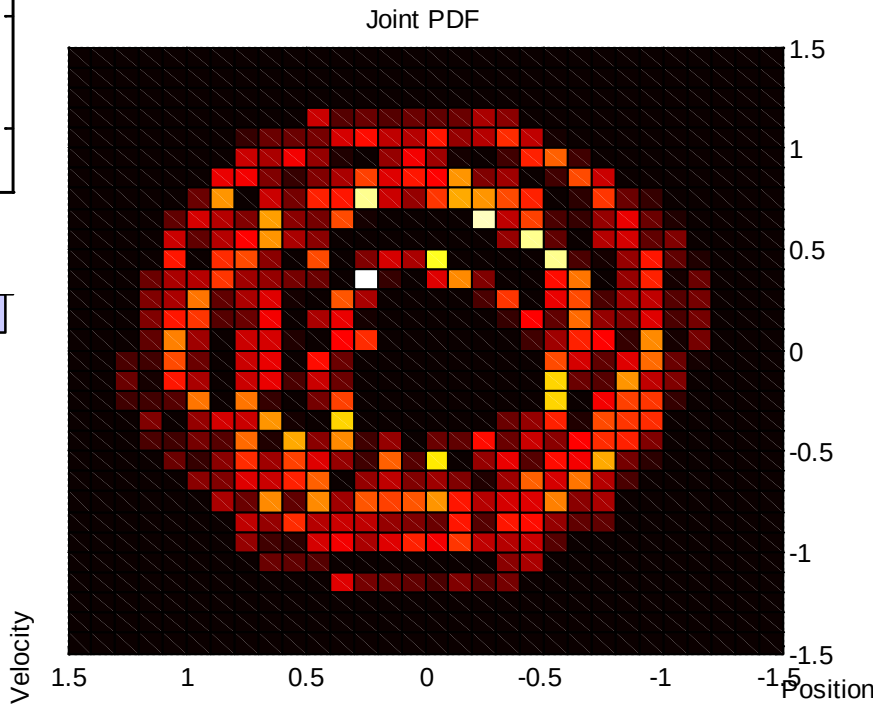




Langevin dynamics







Theory of probability

Motivation

- We need to:
  - to infer about properties/state of a system
  - to predict about outcome of a future event
  - to judge truth of a hypothesis

- Lack of complete certainty
  - in state/properties of the system
  - in outcome of the future event
  - in truth of the hypothesis

- Knowledge of
  - thought experiments
  - repeatability of given experiment, observed data
  - context of problem and similar situations

Three approaches to probability

- **Throw of a fair die.** How likely is a six?
  - Other examples: energy states in Maxwell Boltzmann distribution; fair coin toss; cryptology; lottery design;

- **Throw of a loaded die.** How likely is a six?
  - Other examples: pre-disposition to genetic disease (diabetes, thalassemia, cystic fibrosis etc.); effectiveness of new drug; age-specific mortality; annual maximum wave height; sensitivity & specificity of diagnostic test; psephology

- **Global warming.** How likely is it to be true?
  - Other examples: accuracy of financial model/ weather model/ finite element model/ seismological model etc.; correct location of oil well; sports betting; fairness of a coin/die

Three approaches to probability

- **Classical**: equally likely outcomes – thought experiment
- **Frequentist**: large number of identical trials – actual experiments
- **Judgmental/Bayesian**: degree of belief – use of experience, association, intuition etc.

- **Issues**:
  - Are these three approaches incompatible?
  - Can I mix them to get useful results?

Basic set theory

- Universal set
- Combination of sets
- Set relations
- Countable sets
- Fields
- sigma-fields
- Measure
  - Measurable sets
  - Measure
  - Measure space
  - Measurable function

Set relations

Idempotent laws			
(1*a*)	$A \cup A = A$	(1*b*)	$A \cap A = A$
Associative laws			
(2*a*)	$(A \cup B) \cup C = A \cup (B \cup C)$	(2*b*)	$(A \cap B) \cap C = A \cap (B \cap C)$
Commutative laws			
(3*a*)	$A \cup B = B \cup A$	(3*b*)	$A \cap B = B \cap A$
Distributive laws			
(4*a*)	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	(4*b*)	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
Identity laws			
(5*a*)	$A \cup \varnothing = A$	(5*b*)	$A \cap U = A$
(6*a*)	$A \cup U = U$	(6*b*)	$A \cap \varnothing = \varnothing$
Involution law			
(7)	$(A^c)^c = A$		
Complement laws			
(8*a*)	$A \cup A^c = U$	(8*b*)	$A \cap A^c = \varnothing$
(9*a*)	$U^c = \varnothing$	(9*b*)	$\varnothing^c = U$
DeMorgan`s laws			
(10*a*)	$(A \cup B)^c = A^c \cap B^c$	(10*b*)	$(A \cap B)^c = A^c \cup B^c$

Limits of sets

Recall limit of the sequence of functions $\{f_i(x)\}$. For a given value of x (x is omitted to make the notation cleaner), t he term lim inf f_n denotes the maximum of the sequence of minima:

$$\liminf_{n\to\infty} f_n = \max_{0<j<\infty}\{\min f_j, f_{j+1},\ldots\} = \bigcup_{j=1}^{\infty}\bigcap_{k=j}^{\infty} f_k$$

These concepts directly apply to a sequence of sets $\{A_k\}$. The infimum and supremum of the sequence are the sets defined respectively as:

$$\liminf_{n\to\infty} A_n = \bigcup_{j=1}^{\infty}\bigcap_{k=j}^{\infty} A_k = \left\{w : \sum_k (1 - \mathrm{I}_{A_k}(w)) < \infty\right\} \quad \begin{pmatrix}\text{i.e., set of points absent in a}\\ \text{finite number of } A_k\text{'s}\end{pmatrix} \quad (0.$$

$$\limsup_{n\to\infty} A_n = \bigcap_{j=1}^{\infty}\bigcup_{k=j}^{\infty} A_k = \left\{w : \sum_k I_{A_k}(w) = \infty\right\} \quad \begin{pmatrix}\text{i.e., set of points present in}\\ \text{all } A_k\text{'s}\end{pmatrix} \qquad (0.$$

The limit of the sequence $\{A_k\}$ exists if the two limits are equal and may be denoted A:

$$\liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n = A \qquad (0.$$

which may be written in short as:

$$\lim_{n\to\infty} A_n = A \quad \text{or} \quad A_n \to A \qquad (0.4)$$

Limits of sets: example

$$\text{let } A_k = \left[0, \frac{k}{k+1}\right]$$

$$B_n = \inf_{k \geq n} A_k = \bigcap_{k=n}^{\infty} \left[0, \frac{k}{k+1}\right] = \left[0, \frac{n}{n+1}\right]$$

$$\text{Then, } \lim_{n \to \infty} \inf \left[0, \frac{n}{n+1}\right] = \bigcup_{n=1}^{\infty} \inf_{k \geq n} \left[0, \frac{k}{k+1}\right] = \bigcup_{n=1}^{\infty} \left[0, \frac{n}{n+1}\right] = [0,1]$$

$$B^n = \sup_{k \geq n} A_k = \bigcup_{k=n}^{\infty} A_k = \bigcup_{k=n}^{\infty} \left[0, \frac{k}{k+1}\right] = [0,1]$$

$$\lim_{n \to \infty} \sup A_n = \lim_{n \to \infty} B^n = \lim_{n \to \infty} [0,1] = [0,1]$$

Set algebra

Let X be any set[1]. A non-empty collection $\mathcal{a}$ of subsets of X is an algebra of sets (i.e., a *field*) if: whenever A_1, A_2 are in $\mathcal{a}$, so are $X \backslash A_1$ (i.e., complement of A_1) and $A_1 \cup A_2$ (and therefore $A_1 A_2$ also). Generalizing, if $A_1, A_2, \ldots, A_n$ (n finite) are in $\mathcal{a}$, so are $A_1 \cup A_2 \ldots \cup A_n$ and $A_1 A_2 \ldots A_n$.

Example: Let X={a,b,c}. Then we could define a field $\mathcal{a}$ as:

$\mathcal{a}$ = ϕ, X, {a}, {b,c}

σ algebra (or σ field or Borel field): The algebra described above is a σ algebra of sets if it holds for a countably infinite collection $A_1, A_2, \ldots$. That is, whenever, the sequence $A_1, A_2, \ldots$, belongs to $\mathcal{a}$, so does $\bigcup_{i=1}^{\infty} A_i$. In other words, a σ algebra $\mathcal{a}$ of subsets of a given set X contains the empty set ϕ and is closed with respect to complementation and countable unions.[2]

Measure

Measurable set: A couple (X, $\mathcal{a}$) is a measurable space where X is any set and $\mathcal{a}$ is a σ algebra of subsets of X. A subset A of X is measurable with respect to $\mathcal{a}$ if A ∈ $\mathcal{a}$.

Measure: A measure *m* on a measurable space (X, $\mathcal{a}$) is a non-negative set function defined for all sets of the σ algebra $\mathcal{a}$, if it has the properties:

(i) $m(\phi) = 0.$ (1)

(ii) If $A_1, A_2, \ldots$ is a sequence of disjoint sets of $\mathcal{a}$, then $m(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} m(A_i)$. (2)

Measure space: A measure space (X, $\mathcal{a}$, m) means a measurable space (X, $\mathcal{a}$) together with a measure *m* defined on $\mathcal{a}$.

Measurable function: Let a function *f* be defined on the measurable space (X, $\mathcal{a}$) described above. Let the range of *f* be the extended real line. If *f* satisfies any one of the following conditions, then it is a measurable function with respect to $\mathcal{a}$:

$$\{x : f(x) < \alpha\} \in \mathcal{a} \text{ for each } \alpha$$
$$\{x : f(x) \leq \alpha\} \in \mathcal{a} \text{ for each } \alpha$$
$$\{x : f(x) > \alpha\} \in \mathcal{a} \text{ for each } \alpha$$
$$\{x : f(x) \geq \alpha\} \in \mathcal{a} \text{ for each } \alpha$$

Probability as a Measure - Axioms of probability

A probability space $(\Omega, \mathfrak{F}, P)$

1) $0 \le P(A) \le 1$ for every measurable set $A \in \mathfrak{F}$.

2) $P(\Omega) = 1$

3) If $A_1, A_2 \ldots$ are disjoint sets in $\mathfrak{F}$ then

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

Example: field

Example: throw of a die.

Sample space, $\Omega = \{1,2,3,4,5,6\}$ [More correctly, the sample space should be defined as {obtain 1 dot, obtain 2 dots etc.}, but we have already mapped the sample space on the real line. Also note that we have not included events that the die falls off the table, or stands on a corner etc.]

Depending on your interest, the field $\mathfrak{F}$ can be defined in many ways.

Say, $\mathfrak{F} = 2^{\Omega}$, i.e., the power set of Ω, i.e., all combinations of the 6 sample points.

Or, say, $\mathfrak{F} = \{ \phi, \Omega, \text{odds}, \text{evens}\}$

Or, say, $\mathfrak{F} = \{ \phi, \Omega, 1, \text{evens}, \{1,\text{evens}\}, \{3,5,\text{evens}\}, \{1,3,5\}\}$

In each case, probabilities must be assigned to each member of $\mathfrak{F}$.

Summary – probability theory

- Probability
  - Likelihood of an event
    - Classical definition – thought experiment based
    - Relative frequency based – actual experiment based
    - Bayesian – judgment based
  - P(A) is a fraction between 0 and 1
  - P=0 means impossible event
  - P=1 means sure event
  - If A, B are disjoint events, P(AUB) = P(A) + P(B)

Example

- An urn contains 8 Black, 9 Green and 3 Blue balls. 3 balls are drawn without replacement. Find probability that:
  - All three are black
  - At least one is green
  - One of each colour is drawn

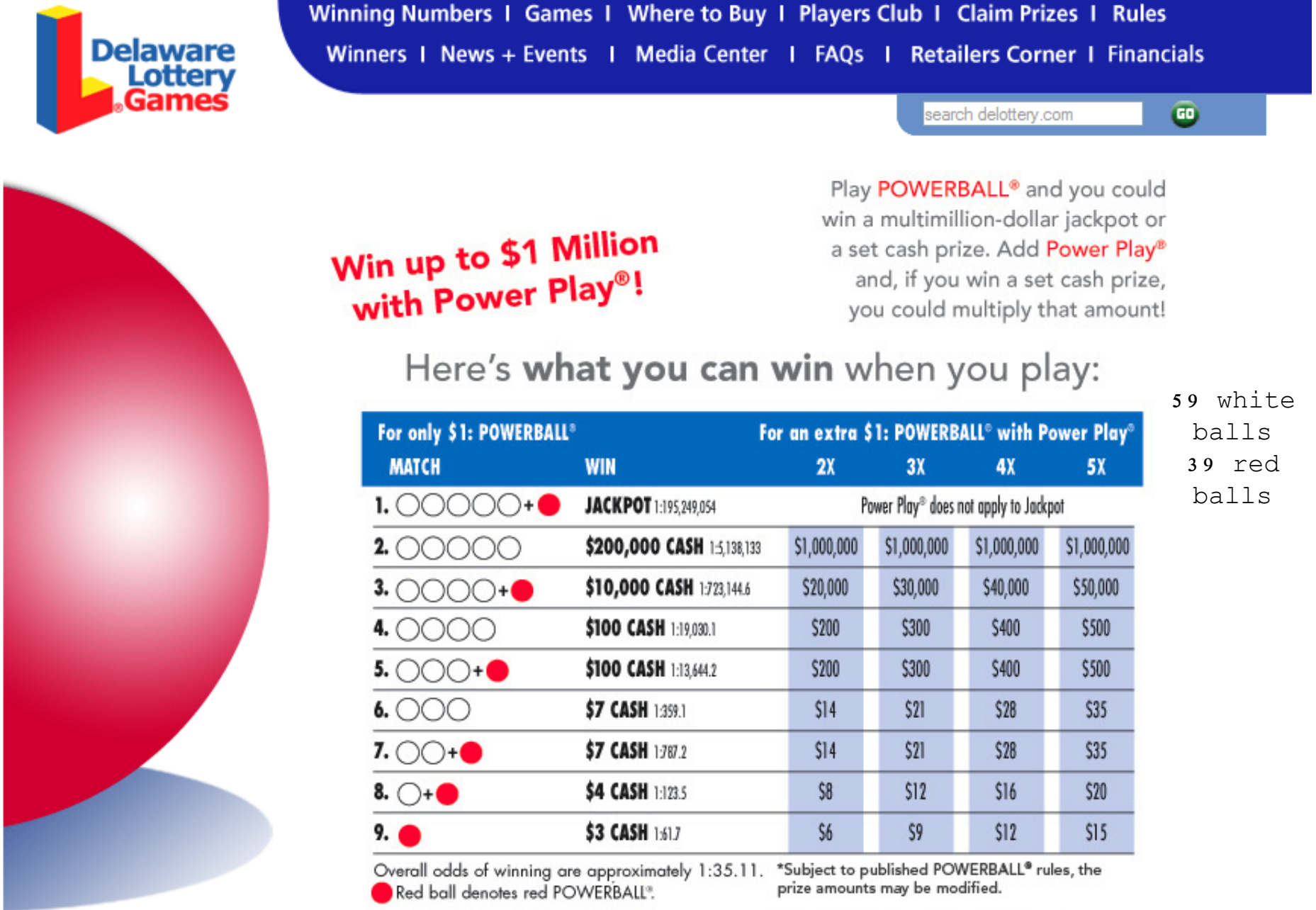
Ans: 0.049, 0.855, 0.189

Example

- Find the probability that in a game of bridge, each player will receive an ace

Ans: 0.105

Lottery design



59 white balls
39 red balls

Example

- *m* boxes and *r* particles, *m>r*
- P[one particle per box in *r* pre-selected boxes]
  - Maxwell Boltzmann Statistics
    - Distinguishable particles
    - No restriction on number of particles per box
    - Number of equally likely ways to place the particles: 1st particle can go to any of m boxes, 2nd particle can go to any of m boxes, …, rth particle can go to any of m boxes. N = m^r
    - Number of favourable ways: 1st particle in r ways, 2nd particle in (r-1) ways, 3rd particle in (r-2) ways, …, rth particle in 1 way. n_A=r!
    - P=*r*!/*m*^*r*

Example

- *m* boxes and *r* particles, *m*>*r*
- P[one particle per box in *r* pre-selected boxes]
  - Bose Einstein Statistics
    - Indistinguishable particles
    - No restriction on number of particles per box
    - Total number of equally likely ways: Two fixed walls on two ends. (m-1) partitions to play with. r particles can go anywhere. Arrange (m-1+r) things of two kinds. N=(m-1+r)!/(m-1)!/r!
    - Number of favourable ways: only one.
    - P=*r*!(*m*-1)!/(*r*+*m*-1)!

Example

- *m* boxes and *r* particles, *m>r*
- P[one particle per box in *r* pre-selected boxes]
  - Fermi Dirac Statistics
    - Indistinguishable particles
    - At most one particle per box
    - Total number of equally likely ways: choose r non-empty boxes from m boxes. N=m!/r!/(m-r)!
    - Number of favourable ways: only one.
    - P=r!(m-r)!/m!

Probability of joint events

Conditional probability

Definition: $P(A \mid B) = \dfrac{P(AB)}{P(B)}$

Joint probability: $P(AB) = P(A \mid B)P(B) = P(B \mid A)P(A)$

Generalizing: $P(A_1 A_2 A_3 \ldots A_{n-1} A_n) =$

$$P(A_n \mid A_{n-1} \ldots A_3 A_2 A_1) P(A_{n-1} \mid A_{n-2} \ldots A_3 A_2 A_1) P(A_3 \mid A_2 A_1) P(A_2 \mid A_1) P(A_1)$$

Statistical independence

Definition: $A \perp B \Leftrightarrow P(A \mid B) = P(A)$

Symmetry: $A \perp B \Leftrightarrow B \perp A$

and $P(A \mid B) = P(A) \Leftrightarrow P(B \mid A) = P(B)$

Mutual independence of n events:

$P(A_{i1} A_{i2} \ldots A_{ir}) = P(A_{i1}) P(A_{i2}) \ldots P(A_{ir}),$

for any r and all subsets $\{i1, i2, \ldots, ir\} \subset \{1, 2, \ldots, n\}$

Total probability

A partition of the sample space:

$$S = \{B_1, B_2, ..., B_n\},$$

such that, $\bigcup_{i=1}^{n} B_i = \Omega$, and $B_i B_j = \varnothing$ for $i \neq j$

For any set, $A = \bigcup_{i=1}^{n} AB_i$

Hence, $P(A) = P(\bigcup_{i=1}^{n} AB_i) = \sum_{i=1}^{n} P(AB_i)$

Can be expanded to: $P(A) = \sum_{i=1}^{n} P(A \mid B_i) P(B_i)$

Bayes' theorem

New knowledge K_1 has been obtained.

Updated probability of hypothesis, B_r :

$$P(B_r \mid K_1) = \frac{P(B_r K_1)}{P(K_1)} = \frac{P(K_1 \mid B_r)P(B_r)}{P(K_1)}$$

Can be expanded to:

$$P(B_r \mid K_1) = \frac{P(K_1 \mid B_r)P(B_r)}{\sum_{i=1}^{n} P(K_1 \mid B_i)P(B_i)}$$

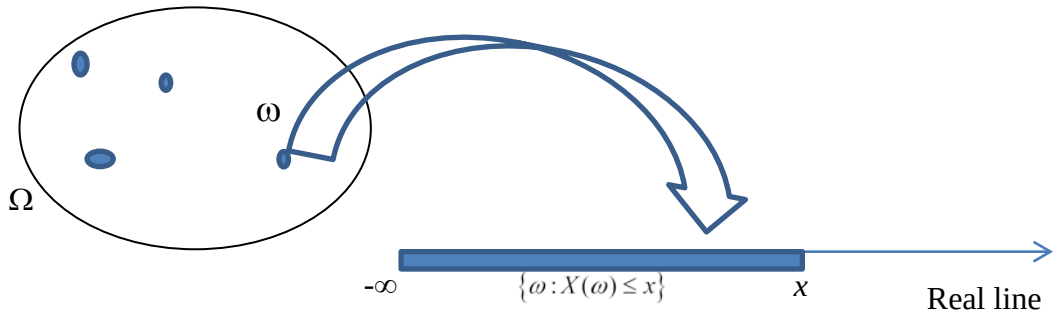
Sequential updating possible

Random variables

Modeling of randomness – random variables

- Random variable
  - numerical values differ from outcome to outcome
  - but shows statistical regularity
  - formally: function defined on sample space
  - defined by: probability distribution type, mean, variance etc.
- Jointly distributed random variables
- Stochastic processes

Random variable



When the possible outcomes of an experiment (or trial) can be given in numerical terms, then we have a random variable in hand. When an experiment is performed, the outcome of the random variable is called a "realization." A random variable can be either discreet, or continuous. A random variable is governed by its probability laws.

If a quantity varies randomly with time, we model it as a stochastic pr ocess. A stochastic process can be viewed as a family of random variables.

If a quantity varies randomly in space, we model it as a random field, which is the generalization of a stochastic process in two or more dimensions.

Formally, a measurable function [1] defined on a sample space is called a random variable (Feller, vol 1, p. 212). That is, X is a random variable if $X = X(\omega)$ is a function defined on the sample space Ω, and for every real x, the set

$$\{\omega : X(\omega) \le x\}$$

is an event in Ω. Thus we confine ourselves to σ-algebra of events of the type $X \le x$. Unless explicitly required, we suppress the argument ω when referring to a random variable in the rest of this text.

Random variables

A random variable is governed by its probability laws. The probabili ty law of a RV can be described by any of the four equivalent ways:

1. CDF (cumulative distribution function)
2. PDF/PMF (probability density function for continuous rv's, probability mass function for discrete rv's)
3. CF (characteristic function)
4. MGF (moment generating function)

- Cumulative distribution function:

The cumulative distribution function of the random variable X is defined as:

$$F_X(x) = P[X \le x] \qquad (0.$$

It starts from 0, ends at 1, and is a non-decreasing function of x. It is piecewise continuous for discrete RVs, and continuous for continuous RVs.

Properties of CDF:

$$F_X(-\infty) = 0$$
$$F_X(\infty) = 1 \qquad (0.$$
$$F_X(x) \text{ is a non-decreasing function of } x$$

Thus, the probability of finding the random variable X in the semi-open interval (a,b] is:

$$P[a < X \le b] = F_X(b) - F_X(a) \qquad (0.3)$$

Expectation

The expectation of any function $g(X)$ of the random variable X is defined as:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)\, f_X(x)dx \quad \text{if X continuous}$$

$$= \sum_{all\, x_i} g(x_i)\, p_X(x_i) \quad \text{if X discrete} \qquad (0.$$

The expectation of a constant is the identity operator:

$$E(c) = c \text{ where } c \text{ is a constant} \qquad (0.$$

Expectation is a linear operator:

$$E(aX + b) = a\, E(X) + b \qquad (0.$$

and if $Y = g_1(X) + g_2(X) + \ldots$, then

$$E(Y) = E(g_1(X)) + E(g_2(X)) + \ldots \qquad (0.4)$$

Expectation

Thus the mean of X is its expectation:

$$\mu = E(X) = \begin{cases} \int_{-\infty}^{\infty} x\, f_X(x)dx, & \text{continuous RV} \\ \sum_{all\, x_i} x_i\, p_X(x_i), & \text{discrete RV} \end{cases} \quad (9.$$

and its variance is the expecation of its squared deviation from the mean:

$$\sigma^2 = E\left[(X-\mu)^2\right] = \begin{cases} \int_{-\infty}^{\infty} (x-\mu)^2\, f_X(x)dx, & \text{continuous RV} \\ \sum_{all\, x_i} (x_i-\mu)^2\, p_X(x_i), & \text{discrete RV} \end{cases}$$

Probability mass function (pmf):

$$p_X(x_i) = P[X = x_i]$$

Such that: $p_X(x_i) \geq 0 \; \forall i$, and $\sum_{\text{all } i} p_X(x_i) = 1$

Cumulative distribution function (cdf):

$$F_X(x_i) = P[X \leq x_i]$$

Mean, $\mu = \sum_{i=1}^{n} p_i x_i$

Variance, $\sigma^2 = \sum_{i=1}^{n} p_i (x_i - \mu)^2$

Common discrete distributions

Common discrete distributions
Prof. B. Bhattacharya, Dept. of Civil Engineering, IIT Kharagpur

Distribution	PMF	CDF	Relation between parameters and moments
Discrete uniform	$p_X(x) = \frac{1}{n},\ x = x_1, x_2, ..., x_n$	Step function of height $1/n$	$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i,$ $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$
Bernoulli	$p_X(x) = px + q(1-x),\ x = 0,1$ where, $p + q = 1$	Steps of height q and p at 0 and 1 respectively.	$\mu = p,$ $\sigma^2 = pq$
Geometric	$p_X(x) = q^{x-1}p,\ x = 1,2,3,....$ where, $p + q = 1$	$F_X(x) = 1 - q^x,\ i = 1,2,3,...$	$\mu = 1/p,$ $\sigma^2 = (1-p)/p^2$
Binomial	$p_X(x) = \binom{n}{x} p^x q^{n-x},\ x = 0,1,2,....,n$ where, $p + q = 1$	Not available in closed form	$\mu = np,$ $\sigma^2 = npq$
Multinomial	$p(x_1, x_2, ..., x_k; p_1, p_2, ..., p_k, n) = \frac{n!}{x_1!x_2!...x_k!} p_1^{x_1} p_2^{x_2} ... p_k^{x_k}$ $\sum_{i=1}^{k} p_i = 1, \sum_{i=1}^{k} x_i = n$	Not available in closed form	$\mu_i = np_i,$ $\sigma_i^2 = np_iq_i$
Negative binomial	$p_X(x) = \binom{x-1}{r-1} p^r q^{x-r},\ x = r, r+1, ...$ where, $p + q = 1$	Can be given in terms of binomial CDF	$\mu = r/p,$ $\sigma^2 = rq/p^2$
Hyper-geometric	$p_X(x) = \frac{\binom{d}{x}\binom{N-d}{n-x}}{\binom{N}{n}},\ x = 0,1,2,....,\min(d,n)$	Not available in closed form	$\mu = nd/N,$ $\sigma^2 = \frac{nd(N-d)}{N^2}\binom{N-n}{N-1}$
Poisson	$p_X(x) = e^{-\mu}\frac{\mu^x}{x!},\ x = 0,1,2,3,....$	Not available in closed form	$\mu = \mu,$ $\sigma^2 = \mu$
Zeta or Pareto	$p_X(x) = \frac{c}{x^{\alpha+1}},\ x = 1,2,3,....,\alpha > 0$ such that $c = 1/\zeta(\alpha+1)$ where $\zeta(s)$ = Reimann zeta fn. $= \sum_{k=1}^{\infty} 1/k^s,\ s > 1$	Not available in closed form	$\mu = \frac{\zeta(\alpha)}{\zeta(\alpha+1)}$

Continuous distributions

Probability density function (pdf):

$$f_X(x)dx = P[x \le X \le x + dx]$$

Such that: $f_X(x) \ge 0 \; \forall x$, and $\int_{-\infty}^{\infty} f_X(x)dx = 1$

Cumulative distribution function (cdf):

$$F_X(x) = P[X \le x]$$

Mean, $\mu = \int_{-\infty}^{\infty} x f_X(x)dx$

Variance, $\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)dx$

Common continuous distributions

Distribution (explanation)	**PDF**	**CDF**	**Relation between parameters and moments**
Uniform	$f_X(x) = \begin{cases} 1/(b-a), & a \le x \le b \\ 0, & \text{otherwise} \end{cases}$	Linearly increases from 0 at a to 1 at b	μ_X = (b+a)/2 $\sigma_X^2 = (b-a)^2/12$
Normal	$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right)$, $-\infty < x < \infty$	Not available in closed form. Can be given in terms of the standard normal CDF, Φ: $F_X(x) = \Phi\left(\frac{x-\mu_X}{\sigma_X}\right)$	Obvious
Gamma	$f_X(x) = \lambda\frac{(\lambda x)^{k-1}}{\Gamma(k)}e^{-\lambda x}$, $x > 0$ where $\Gamma(k) =$ gamma fn.$= \int_0^\infty t^{k-1}e^{-t}dt$ k any positive real number	$F_X(x) = \frac{\Gamma(\lambda x, k)}{\Gamma(k)}$ where $\Gamma(x,\alpha) =$ incomplete gamma fn$= \int_0^x e^{-t}t^{\alpha-1}dt$	$\mu_X = k/\lambda$ $\sigma_X^2 = k/\lambda^2$
Chi-squared with n dof (sum of n independent squared standard normal variables) (Gamma with k=n/2 and λ=1/2)	$f_X(x) = \frac{1}{2^{n/2}\Gamma(n/2)}e^{-x/2}x^{n/2-1}, x \ge 0$ n does not have to be integer	$F_X(x) = \frac{\Gamma(x/2, n/2)}{\Gamma(n/2)}$, $x > 0$	$\mu_X = n$ $\sigma_{X^2} = 2n$
Chi with n dof (square root of Chi-squared random variable with dof n) Chi with n= 1 is called "half normal", with n = 2 is Rayleigh, and n=3 is MB	$f_X(x) = \frac{1}{2^{n/2-1}\Gamma(n/2)}e^{-x^2/2}x^{n-1}, x \ge 0$	$F_X(x) = \frac{\Gamma(x^2/2, n/2)}{\Gamma(n/2)}$, $x > 0$	$\mu_X = \sqrt{2}\frac{\Gamma((n+1)/2)}{\Gamma(n/2)}$ $\sigma_X^2 = n - \mu_X^2$
Student's t distribution (ratio of standard normal to chi with dof n)	$\frac{\Gamma((n+1)/2)}{\sqrt{\pi n}\Gamma(n/2)}(1+x^2/n)^{-(n+1)/2}, -\infty < x < \infty$		$\mu = 0$ $\sigma^2 = \frac{n}{n-2}, n > 2$
F distribution (ratio of two chi-squared random variables with dofs m and n)	$\frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)}\left(\frac{m}{n}\right)^{m/2}x^{m/2-1}\left(1+\frac{mx}{n}\right)^{-(m+n)/2}, x > 0$		$\mu = \frac{n}{n-2}, n > 2$ $\sigma^2 = \frac{n^2(2m+2n-4)}{m(n-2)^2(n-4)}, n > 4$

Common continuous distributions

Distribution (explanation)	PDF	CDF	Relation between parameters and moments
Lognormal (exponentiated normal)	$f_X(x) = \frac{1}{\sqrt{2\pi}\zeta x}\exp\left[-\frac{1}{2}\left(\frac{\ln x - \lambda}{\zeta}\right)^2\right], \quad x > 0$	Not available in closed form. Can be given in terms of the standard normal CDF, Φ : $F_X(x) = \Phi\left(\frac{\ln x - \lambda}{\zeta}\right)$	$\zeta = \sqrt{\ln(1+V_X^2)}$ $\lambda = \ln(\mu_X) - \frac{1}{2}\zeta^2$
Maxwell Boltzmann (Chi with n=3)	$f_X(x) = \frac{\sqrt{2}}{\sqrt{\pi}} e^{-x^2/2} x^2, x \geq 0$	$F_X(x) = \frac{\Gamma(x^2/2, 3/2)}{\sqrt{\pi}/2}, \quad x > 0$	$\mu_X = \sqrt{2}\frac{2}{\sqrt{\pi}}$ $\sigma_X^2 = 3 - \frac{8}{\pi}$
Gumbel (for maxima)	$f_X(x) = \alpha e^{-\alpha(x-u)} e^{-e^{-\alpha(x-u)}}$, $-\infty < x < \infty$	$F_X(x) = e^{-e^{-\alpha(x-u)}}$, $-\infty < x < \infty$	$\sigma_X = \frac{\pi}{\sqrt{6}\alpha}$ $\mu_X = u + \frac{0.5772}{\alpha}$
Frechet (for maxima)	$f_X(x) = \alpha k (x-\lambda)^{-k-1} e^{-\alpha(x-\lambda)^{-k}}$ $k, \alpha > 0, x > \lambda$	$F_X(x) = e^{-\alpha(x-\lambda)^{-k}}$ $k, \alpha > 0, x > \lambda$	$\mu = \lambda + \alpha^k \Gamma(1 - 1/k)$ $\sigma^2 = \alpha^{2k}\left[\Gamma(1-2/k) - \Gamma^2(1-1/k)\right]$
Weibull (for minima)	$f_X(x) = \frac{k}{u}\left(\frac{x-x_0}{u}\right)^{k-1} \exp\left(-\left(\frac{x-x_0}{u}\right)^k\right),$ $x > x_0$	$F_X(x) = 1 - \exp\left(-\left(\frac{x-x_0}{u}\right)^k\right)$ $x > x_0$	$\mu = x_0 + u\,\Gamma(1 + 1/k)$ $\sigma^2 = u^2\left[\Gamma(1+2/k) - \Gamma^2(1+1/k)\right]$
Exponential (Weibull with x_0=0, k=1) (Gamma with k=1)	$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0$	$F_X(x) = 1 - e^{-\lambda x}, \quad x \geq 0$	$\lambda = 1/\mu_X$ $\sigma_X = 1/\lambda$
Rayleigh (Weibull with x_0=0, k=2) (Chi with 2 dofs)	$f_X(x) = \frac{2}{u}\left(\frac{x}{u}\right) \exp\left(-\left(\frac{x}{u}\right)^2\right)$		
Wald (inverse Gaussian) (time taken by a Brownian particle to reach distance *d* for the first time with drift velocity *v* and diffusion coefficient β. Here $\mu = d/v, \lambda = d^2/\beta$)	$f_X(x) = \left(\frac{\lambda}{2\pi x^3}\right)^{\frac{1}{2}} \exp\left[-\frac{\lambda}{2\mu^2}\frac{(x-\mu)^2}{x}\right], x > 0$		$\mu_X = \mu$ $\sigma_X{}^2 = \mu^3/\lambda$

Common continuous distributions

Distribution (explanation)	PDF	CDF	Relation between parameters and moments
Cauchy	$f_X(x) = \frac{1}{\pi\lambda}\left[1+(\frac{x-\theta}{\lambda})^2\right]^{-1}, \lambda > 0$ $\theta =$ location parameter, $\lambda =$ scale parameter	$F_X(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\frac{x-\theta}{\lambda}$	μ_X does not exist. No finite moment of order 1 or greater exists
Beta	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1},$ $0 < x < 1, \alpha > 0, \beta > 0$		$\mu = \frac{\alpha}{\alpha+\beta}$ $\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
Pareto ("heavy tailed")	$f_X(x) = kk_0x^{-k}$ $x > x_0$	$F_X(x) = 1 - k_0x^{-k}$ $k_0, k > 0,$ $x > x_0 = \frac{1}{k_0^k} > 0$	

Random number generation

Random Number Generation for MCS

- Monte-Carlo simulation is a method of solving numerical problems by generating a series of random variables
- Series of (pseudo)-random numbers are computer-generated
- Properties of a good generator:
  - Accuracy
  - Long Period
  - High Speed
  - Short Setup time
  - Small Length of compiled code
  - Machine independence
  - Versatility in possible applications
  - Simplicity and readability

Monte-Carlo Simulation - Application

- Establish statistical properties of functions of random variables/processes (no closed-form solution)
- Estimate probabilities of rare events
- Obtain evolution or response of stochastic systems
- Evaluation of definite integrals and expectations

Simulation of Random Numbers

- multiplicative congruential generator:

$$x_i = cx_{i-1} \mod m$$

$$u_i = x_i / m$$

- Seed: x_0
- Period of generator, m_0: $x_{i+m_0} = x_i$
- Generally: $m \sim 2^b$ $\quad b = \text{bits/word}$

$$c \sim \sqrt{m}$$

$$m_0 \sim 2^{b-2}$$

"Best" RN Generator:

$m = 2^{31} - 1$

$c = 7^5 = 16807$

$m_0 = m$ (full period)

Example: RN generation

$$x_i = cx_{i-1} \mod m$$

$$u_i = x_i / m$$

Given: $c = 13,\ m = 100$

set: $x_0 = 1$

Sequence: 1, 13, 69, 97, 61, 93, 9, 17, 21, 73,
49, 37, 81, 53, 57, 41, 33, 29, 77

Period = 20

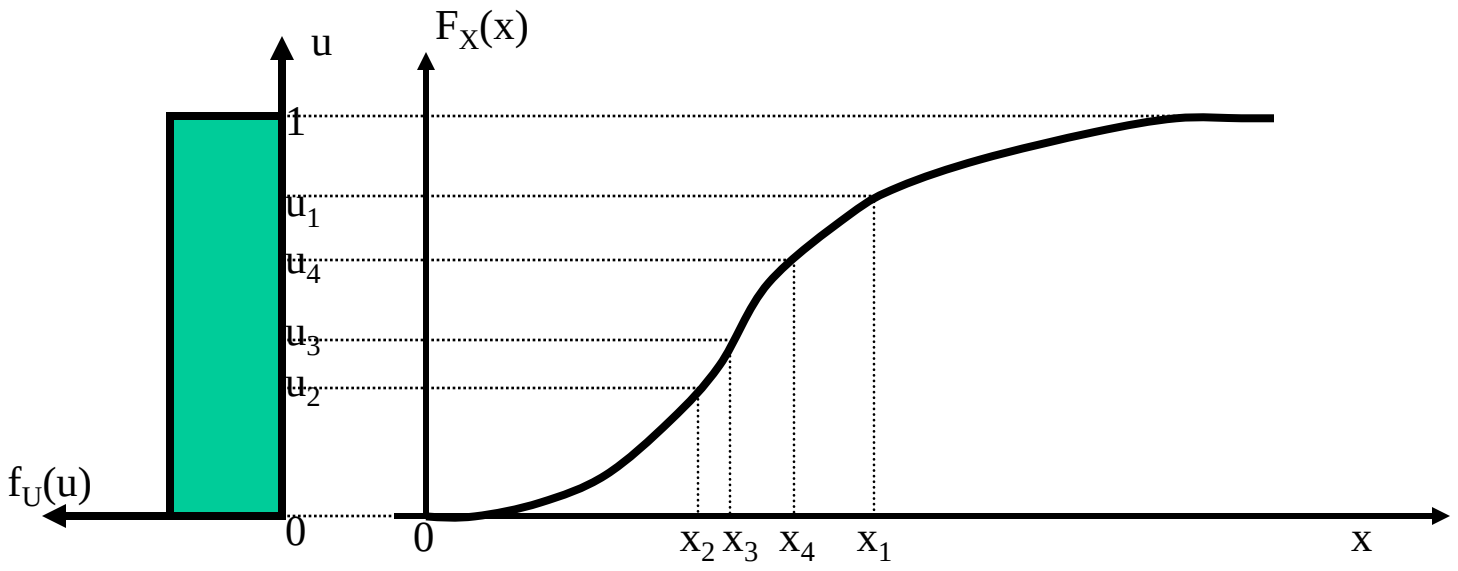
Generating Continuous Random Variables

Theorem: If F is a continuous CDF,
and if U is a Uniform RV in (0,1),
then the RV X defined as

$$X = F^{-1}(U)$$

is distributed according to F.

Inversion method



Example of Inversion Method

- Example: Generate X~Exp(λ)
- Generate u~U(0,1)
- Invert: $F_X(x)=1 – \exp(-\lambda x) = u$
- Hence, x = - (1/l)ln(1- u)
- Return x

Inversion Method for Discrete RVs

- Generate U(0,1) as before

$$x = F_X^{-1}(u) = \sup\{x_k : F_X(x_k \geq u, k = 1, 2, \ldots\}$$

$$\text{i.e., set } X = x_k$$

$$\text{iff } p_1 + p_2 + \ldots + p_{k-1} < u$$

$$\leq p_1 + p_2 + \ldots + p_{k-1} + p_k$$

Example of Discrete RV Generation using Inversion

- Example: N~Geometric(p)
- q = 1 – p
- $F_N(n)= 1- q^n$

- generate u ~ U(0,1)
- n = sup { x: $1 – q^x >= u$, x integer}

  [Equivalently, n = int (ln(1-u)/ln(q) + 1]
- return n

Generating uncorrelated normals

1.1.1.1 Generating independent normal vectors

A vector of uncorrelated normals, $\underline{X}$ (with mean vector $\underline{\mu}$ and covariance matrix $\underline{V} = \sigma_i^2 \delta_{ij}$) can be generated:

(i) By inverting the standard normal distribution function Φ:

- Generate independent uniforms $\underline{U}$, $U_i \sim U(0,1)$:
- Invert each of them according to:

$$x_i = \mu_i + \sigma_i \Phi^{-1}(u_i) \qquad (0.$$

- return $\underline{x}$

(ii) By generate unit vectors in standard normal space, and generate radial distance.

- Generate *n* independent angles, $\Theta_i \sim U(0, 2\pi)$
- generate squared radius R^2 (independent of the vector $\underline{\Theta}$), $R^2 \sim \chi^2(n)$ is chi-squared with d.o.f. *n*. Then, each X_i is given by:

$$x_i = \mu_i + \sigma_i r \cos(\theta_i) \qquad (0.$$

- return $\underline{x}$

Generating correlated normals

1.1.1.1 Generating correlated normals

A vector of correlated Normals, $\underline{X}$, (with mean vector $\underline{\mu}$ and covariance matrix $\underline{V}$) can be generated as follows:

- Perform Cholesky fa ctorization of the covariance matrix: let $\underline{C}$ be the lower triangle factor of $\underline{V}$, i.e., $\underline{C}\underline{C}' = \underline{V}$
- Generate independent standard normals $\underline{z}$ as above.
- Convert them according to

$$\underline{x} = \underline{\mu} + \underline{C}\underline{z} \qquad (0.$$

- return $\underline{x}$

Joint probability densities

Jointly distributed random variables

We start with the joint cumulative distribution function (JCDF) of two random variables X and Y. It is given by the probability:

$$F_{X,Y}(x,y) = P[X \le x, Y \le y] \qquad (9.$$

It must be a monotone function taking values between 0 and 1. In the discrete case, it is given by the sum of the joint probabiliy mass function (JPMF):

$$F_{X,Y}(x,y) = P[X \le x, Y \le y] = \sum_{y_j \le y} \sum_{x_i \le x} p_{X,Y}(x_i, y_j) \qquad (9.$$

while in the continuous case, it is given by the integration of the joint probability density function (JPDF), $f_{X,Y}(x,y)$,

$$F_{X,Y}(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u,v)\,du\,dv \qquad (0.3)$$

The JPMF is a non -negative function and sums to one . Its interpretation of the JPMF is as in the one variable case:

$$p_{X,Y}(x_i, y_j) = P[X = x_i, Y = y_j] \qquad (9.$$

Likewise, the joint probability density function of two continuous random variables non-negative, contains a volume of unity under it, and is interpreted as:

$$f_{x,y}(x, y)\ \ \Delta x \Delta y = P[X \in (x, x + \Delta x) \bigcap Y \in (y, y + \Delta y)] \qquad (0.2)$$

The probability content of a region A can be given by:

$$P[(x, y) \in A] = \begin{cases} \sum_{all\ y} \sum_{all\ x} I_A p_{X,Y}(x, y), \text{ discrete case} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I_A f_{X,Y}(x, y) dx\, dy, \text{continuous case} \end{cases} \qquad (0.$$

where $I_A = \begin{cases} 1, & \text{if } (x, y) \in A \\ 0, & \text{otherwise} \end{cases}$

The conditional PMF of X given Y has taken a particular value is:

$$p_{X|Y=y}(x,y) = P[X = x \,|\, Y = y] = \frac{p_{X,Y}(x,y)}{p_Y(y)} \qquad (0.$$

The JPMF can be written in terms of the conditional and marginal PMFs as:

$$p_{X,Y}(x,y) = p_{X|Y=y}(x,y)\, p_Y(y) = p_{Y|X=x}(y,x)\, p_X(x)$$

The conditional PDF of X given a particular realization of Y is,

$$f_{X/Y=y}(x,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \qquad (0.$$

The explanation can be given as:

$$\begin{aligned} f_{X|Y=y}(x,y)dx &\approx P[X \in (x, x+dx) \,|\, Y \in (y, y+dy)] \\ &= \frac{P[X \in (x, x+dx), y \in (y, y+dy)]}{P[Y \in (y, y+dy)]} \\ &= \frac{f_{X,Y}(x,y)dxdy}{f_Y(y)dy} \end{aligned}$$

The joint PDF can be given as the product of the conditonal and the corresponding marginal:

$$f_{X,Y}(x,y) = f_{X|Y=y}(x,y)\, f_Y(y) = f_{Y|X=x}(y,x)\, f_X(x) \qquad (0.2)$$

Linear combination of random variables

Let Y be a linear combination of *n* random variablex, $\underline{X}$.

$$Y = a_0 + \underline{A}^T \underline{X}$$

where $\underline{A}^T = \{a_1, a_2, ..., a_n\}$. Let $\underline{\mu}_X$ be the mean vector of $\underline{X}$ and let

$$\underline{V}_X = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ \sigma_{n1} & \cdots & \cdots & \sigma_{nn} \end{bmatrix}$$

be the covariance matrix of $\underline{X}$, so that $\sigma_{ij} = \mathrm{cov}(X_i, X_j)$. Then the mean of *Y* is

$$\mu_Y = a_0 + \underline{A}^T \underline{\mu}_X$$

and the variance of Y is:

$$\sigma_Y{}^2 = \underline{A}^T \underline{V}_X \underline{A}$$

Further, if $\underline{X}$ is jointly Normal (discussed next), *Y* is Normal too.

Convergence of random variables

A sequence[1] of real numbers is c alled convergent if it has a limit. A real number l is the limit of a sequence if for each positive ε there is an N such that for all $n \geq N$ we have $|x_n - l| < \varepsilon$. A sequence can have at most one limit, and conventionally, +/ - ∞ is not considered a valid limit. Also, a sequence is convergent if and only if it is a Cauchy sequence. A sequence is a Cauchy sequence if given $\varepsilon > 0$ there is an N such that for all n≥N and all m≥N, we have $|x_n - x_m| < \varepsilon$.

Now consider a sequence of random variables $\{X_1, X_2, ..., X_n\}$. Not all sequences or random variables converge to anything. But in some cases we know they do, as in the mean of n iid random variables. Can we generalized this? The question whether a sequence of RVs converge arises naturally in cases of differentiation and integration of a stochastic process, $X(t)$.

Define the "derivative" of process X the usual way as:

$$Y(t,h) = \frac{X(t+h) - X(t)}{h},\ h \to 0 \qquad (0.$$

What does it mean? Is Y a legitimate stochastic process? If so, in what sense? What and how does it converge to?

Consider a probability space (Ω, $\mathfrak{F}$, P). The statement that two random variables X' and X are equal "almost surely" means that, except for events belonging to a set of zero measure, the statement is true with probability 1:

$$X' = X \text{ a.s.} \Leftrightarrow P[X'(w) = X(w)] = 1 \quad \forall w \in A \qquad (0.$$

where $A, A^c \subset B$ and $P(A^c) = 0$. The set A^c is called the exception set.

This brings us to the definition of almost sure (a.s.) c onvergence of a sequence of random variables:

$$\lim_{n\to\infty} P[X_n = X] = 1 \qquad (0.$$

This is spoken as "a.s." convergence and written as $X_n \xrightarrow{a.s.} X$. Eq (0.2) is equivalent to any of the following statements:

(i) $X_n \to X$ except for a set of events with probability zero. That is, there is a measurable set A such that P(A)=1and for every w in A, $\lim X_n(w)=X(w)$.

(ii) $$\lim_{n\to\infty} P\left[\bigcup_{k\geq n} | X_n - X | \geq \varepsilon\right] = 0 \qquad (0.3)$$

(iii) $$\limsup_{n\to\infty} X_n(w) = \liminf_{n\to\infty} X_n(w) \text{ for "almost all" } w \qquad (0.$$

For a.s. convergence to be relevant, all RVs must be defined on the same sample space. Further, such RVs are generally highly dependent.

1.1.1 Convergence in Lp

$$\lim_{n\to\infty} E[|X_n - X|^p] = 0 \qquad (0.$$

When p = 2, this is mean square convergence.

Mean square convergence is also written as $\text{l.i.m.}_{x\to\infty} X_n = X$ and pronounced as "limit in mean."

Mean square convergence is meaningful only if the random variable is second -order, i.e., $E[X_i^2] < \infty$ for every *n*.

1.1.1 Convergence in probability

For any ε > 0,

$$\lim_{n\to\infty} P[|X_n - X| > \varepsilon] = 0 \qquad (0.1)$$

1.1.1 Convergence in distribution

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \quad \text{for every continuity point } x \text{ of } F_X.$$

It is the weakest mode of convergence. If a sequence of RVs is defined in terms of a sequence of the parameter (plural?), then convergence of the parameter sequence usually means convergence in distribution for the sequence (with the limiting value of the parameter).

1.1 Law of large numbers

Consider the case of partial sum S_n of n RVs that are mutually independent, but not necessarily identiacally distributed:

$$S_n = X_1 + X_2 + ... + X_n \qquad (0.$$

Let $\mu_k = E(X_k)$, $\sigma_k^2 = \mathrm{var}(X_k)$ if they exist. Define Y_n=S_n/n , i.e, the average. Does the sequence Y_n converge to anything? Say the se quence Y_n "converges" in some sense to μ. The nature of convergence will depend on the probability structure of the X_i's. And the nature of convergence determines which Law of Large Numbers governs – the strong type or the weak type.

Strong law: If the co nvergence is in L_2 norm, or almost surely, then we have the strong law of large numbers. The convergence of the series $\Sigma\sigma_k^2 / k^2$ is a sufficient condition for the Strong Law to hold for the sequence of mutually independent RVs (Kolmogorov criterion). Also, if the sequence is IID and the mean exists, the Strong Law holds.

Weak law: If the convergence is only in probability, then we have the weak law of large numbers. The Weak Law holds whenever the Xk are uniformly bounded, i.e., wheneve r there exists a constant A such that $|X_k| < A$ for all k. Another sufficient condition for the Weak Law to hold is $(1/n^2)\Sigma\sigma_k^2 \to 0$.

Statistical independence

If X and Y are independent, the conditional distribution (or density or mass) of one is identical to its ma rginal. Equivalently, the joint distribution (or density or mass) is the product of the marginals. Each of these is both a necessary and sufficient condition for independence of X and Y.

$$X \text{ is independent of } Y \Leftrightarrow \begin{cases} F_{X|Y=y}(x,y) = F_X(x) \text{ for all } x, y \text{ (continuous or discrete)} \\ F_{X,Y}(x,y) = F_X(x)F_Y(y) \text{ for all } x, y \text{ (continuous or discrete)} \\ p_{X|Y=y}(x,y) = p_X(x) \text{ for all } x, y \text{ (discrete)} \\ f_{X,Y}(x,y) = f_X(x)\, f_Y(y) \text{ for all } x, y \text{ (continuous)} \end{cases}$$

(0.1)

If X and Y are independent, so are $g(X)$ and $h(Y)$

1.1.1.1 Generalization to n dimensions

A vector $\{X_t, t \epsilon T\}$ of random variables is mutually independent iff for all subsets J of T, the joint CDF is the product of the marginal CDFs:

$$F_J(x_t, t \in J) = \prod_{t \in J} P[X_t \le x_t] \quad (0.1)$$

Joint normal

Recall that X is said to have a normal distribution with mean μ and variance $\sigma^2 > 0$ if:

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma^{1/2}} \exp[-\frac{1}{2} Q_1(x;\mu,\sigma^2)] \qquad (0.$$

where, $Q_1(x;\mu,\sigma^2) = \frac{1}{\sigma^2}(x-\mu)^2 = (x-\mu)\sigma^{-2}(x-\mu)$

Aside : $f(x;\mu,\sigma^2) = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)$

In a parallel manner, a two dimensional RV $\underline{X} = (X_1, X_2)^T$ is said to have a non-singular bivariate normal distribution if its density function is of the form:

$$f(\underline{x};\underline{\mu},\underline{\underline{V}}) = \frac{1}{2\pi|V|^{1/2}} \exp[-\frac{1}{2} Q(\underline{x};\underline{\mu},\underline{\underline{V}})] \qquad (0.$$

$$Q_2(\underline{x};\underline{\mu},\underline{\underline{V}}) = (\underline{x}-\underline{\mu})^T \underline{\underline{V}}^{-1}(\underline{x}-\underline{\mu})$$

where $\underline{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \underline{\underline{V}} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$

$\sigma_i^2 > 0, |\sigma_{12}| < \sigma_1\sigma_2$

Joint normal

An *n* dimensional random variable $\underline{X}$ with mean $\underline{\mu}$ and cov. matrix $\underline{V}$ is said to have a nonsingular multivariate normal distribution if $\underline{V}$ is positive definite and the joint PDF of $\underline{X}$ is :

$$f_{\underline{X}}(\underline{x};\underline{\mu},\underline{V}) = \frac{1}{(2\pi)^{n/2}\,|\underline{V}|^{1/2}} \exp\left(-\frac{1}{2}(\underline{x}-\underline{\mu})^T \underline{V}^{-1}(\underline{x}-\underline{\mu})\right), |V| > 0 \qquad (9.$$

which is symbolically written as $\underline{X} \sim N_n(\underline{\mu},\underline{V})$.

Stochastic processes

Stochastic processes

A stochastic process (also called random function) is a family of random variables $\{X(t), t \in T\}$ indexed by a parameter (or index) t belonging to an index set T. Heuristically it is a function of "time", whose time histories (or sample functions, or realizations) are generally different in different trials, but follow the probability law governing the process. The probability law governing the process is most generally described by a finite dimensional joint probability distribution.

The parameter t can be discrete or continuous. The range of $X(t)$ for a given t, may be discrete or continuous. If the index set is finite, the stochastic process is a random vector, if T is countably infinite, we have a random sequence. For a particular $t_1, X_1 = X_{t_1} = X(t_1)$ is a random variable, and x_{t_1} is one realization of X_{t_1}. Likewise, $x(t)$ is one sample function of $X(t), t \in T$.

If the index s et is finite, the SP is a random vector, if T is countably infinite, we have a random sequence.

For particular $t_1, X_1 = X_{t_1} = X(t_1)$ is a random variable

x_{t_1} is one realization of X_{t_1}

$x(t)$ is one sample fn. of X(t) $\quad t \in T$

Thus $X(t)$ represents an "emsemble" of sample functions.

In general, the probability law $f_{X(t_1),X(t_2)....,X(t_n)}(x_1, x_2....., x_n)$ for the SP $X(t)$ is described by the joint pdf of $X(t_1), X(t_2), ..., X(t_n)$ for all integers n, and for all instances $(t_1, t_2, ..., t_n) \in T$.

Which means we need :

$f_{X(t1)}(x_1)$ for all t_1

$f_{X(t_1),X(t_2)}(x_1, x_2)$ for all t_1, t_2

$\vdots$

$f_{X(t_1),X(t_2),....,X(t_n)}(x_1, x_2, ..., x_n)$ for all $t_1, t_2, ..., t_n$

$\vdots$

so on.

Of course if we know $f_{X(t_1),X(t_2)}(x_1, x_2)$, we also know $f_{X(t_1)}(x_1)$.

1.1.1.1 Mean value function

$$\mu_X(t) = E\left(X\left(t\right)\right) = \int_{-\infty}^{\infty} x f_{X(t)}(x)dx. \qquad (0.$$

1.1.1.2 Variance function

$$\sigma_X^2(t) = \int_{-\infty}^{\infty} \left[x - \mu_X(t)\right]^2 f_{X(t)}(x)dx \qquad (0.$$

1.1.1.3 Auto covariance function (ACVF)

$$\begin{aligned} K_{XX}(s,t) &= \operatorname{cov}\left[X\left(s\right), X\left(t\right)\right] = E\left(X(t)X(s)\right) - E\left(X(t)\right)E\left(X(s)\right) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(x_1 - \mu_X(s)\right)\left(x_2 - \mu_X(t)\right) f_{X_1(s),X(t)}(x_1, x_2)dx_1dx_2 \end{aligned} \qquad (0.$$

$\therefore \sigma_{X(t)}^2 = K_{XX}(t,t)$

1.1.1.4 Autocorrelation function (ACF)

$$\phi_{XX}(s,t) = E\left[X(s)X(t)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_2 f_{X(s),X(t)}(x_1, x_2)dx_1dx_2 \qquad (0.$$

so that the autocovariance function is:

$$K_{XX}(s,t) = \phi_{XX}(s,t) - \mu_X(s)\mu_X(t) \qquad (0.5)$$

Stationarity

Stationary processes – weak and strong stationarity

It is very convenient to separate SPs into two broad classes: stationary and non-stationary (i.e. evolutionary).

Stationarity involves some kind invariance; here it is w.r.t. time.

Let us approach stationarity step by step.

X(t) is "mean-value stationary" if the mean value fn is invariant under a time shift. i.e,

$$\mu_X(t+r) = \mu_X(t) \quad \text{for any } r \qquad (0.$$

This is true only if $\mu_X(t) = \mu_X$ is independent of t.

1.1.1 Weak stationarity

X(t) is "covariance stationary" (or second moment stationary or weakly stationary or simply "stationary") if

$$K_{XX}(s+r,t+r) = K_{XX}(s,t) \text{ for any } r \qquad (0.$$

or, equivalently,

$$\phi_{XX}(s+r,t+r) = \phi_{XX}(s,t) \text{ for any } r \qquad (0.$$

This is possible only if
$K_{XX}(s+r,t+r)$ is a function of $(t-s)$ and not of t or s individually.

$$\begin{aligned} \therefore K_{XX}(s,t) &= G(t-s) \text{ for a covariance stationary process} \\ &= G(\tau) \end{aligned} \qquad (0.$$

where $\tau = t - s =$ time shift $=$ lag . Equivalently, the process X(t) is covariance stationary, if

$$\phi_{XX}(s,t) = R(t-s) = R(\tau) \qquad (0.$$

This immediately means the variance function $\sigma_X^2(t) = \sigma_X{}^2$ is independent of t.

A process X(t) is "stationary of order k" if $\{X(t_1), X(t_2), ..., X(t_k)\}$ and $\{X(t_1+r), X(t_2+r), ..., X(t_k+r)\}$ are identically distributed for an arbitrary set of k points $t_1, ..., t_k$ and any r:

$$f_{X(t_1),X(t_2),.....,X(t_k)}(x_1, x_2, ..., x_k) = f_{X(t_1+r),X(t_2+r),.....,X(t_k+r)}(x_1, x_2, ..., x_k) \quad (0.$$

A process X(t) is "strictly stationary" if above holds for any k and any r.

In particular, second order stationarity is written as:

$$f_{X(t_1),X(t_2)}(x_1, x_2) = f_{X(t_1+r),X(t_2+r)}(x_1, x_2) \text{ for any } t_1, t_2, r \quad (0.$$

and first order stationarity is written as:

$$f_{X(t)}(x) = f_{X(t+r)}(x) \text{ for any } t, r \quad (0.$$

which means that it is independent of time.

If $X(t)$ is k^{th} order stationary, then it is 1^{st} order stationary, 2^{nd} order stationary, …., upto $(k-1)^{th}$ order stationary.

k^{th} moment stationarity may be defined similarly for any $t_1, ..., t_k$ and any r :

$$E[X(t_1)X(t_2)...X(t_k)] = E[X(t_1+r)X(t_2+r)...X(t_k+r)] \qquad (0.$$

It should be obvious that k^{th} order stationarity is a stronger statement than k^{th} moment stationarity.

k^{th} order stationary $\Rightarrow$ k^{th} moment stationary, but the converse is not necessarily true.

Further, k^{th} moment stationary does not imply $(k-1)^{\text{th}}$ moment stationary.

Of course verifying strict stationarity is very tedious & cumbersome if at all possible. So we remain satisfied with the much more lax condition of "second order stationarity" or "weak stationarity" described above. It helps in many many practical situations. Just as we most often get estimates of only μ, σ for a random variable, for a SP too we most often get mean value and covariance functions only.

One notable exception is the Gaussian process which is completely defined by its mean and covariance functions. Therefore, for a Gaussian process, covariance stationarity and strict stationarity are equivalent conditions. This is the reason that Gaussian processes hold such a prominent place in the canons of stochastic processes.

Power spectrum

1.1.1.1 Power spectral density

F.T. does not exist for a weakly stationary random process, because (almost all) x(t)'s are not absolutely integrable in $(-\infty,\infty)$, because x(t) dose not vanish at $-\infty$ and ∞.

But the F.T. of the ACVF (G_{XX}) of a weakly stationary process does exist (in most cases). Which gives the power spectrum of X(t):

$$S_{XX}(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}G_{XX}(\tau)e^{-i\omega\tau}d\tau \quad =\frac{1}{\pi}\int_{0}^{\infty}G_{XX}(\tau)e^{-i\omega\tau}d\tau \qquad (0.$$

and the inverse FT gives back:

$$G_{XX}(\tau)=\int_{-\infty}^{\infty}S_{XX}(\omega)e^{i\omega\tau}d\omega \qquad (0.$$

Requirment: $G_{XX}(\tau)$ is absolutely integrable:

$\int_{-\infty}^{\infty}\left|G_{XX}(\tau)\right|d\tau<\infty,$ does not have a purely periodic component

Properties: $S_{XX}(\omega)\geq 0$ *for all* ω

$S_{XX}(\omega)$ *is real for all* ω

$S_{XX}(\omega)=S_{XX}(-\omega)$ *for all* ω.

Other relations: $G_{XX}(0)=\sigma_X{}^2=\int_{-\infty}^{\infty}S_{XX}(\omega)d\omega.$

you can do all of above by replacing G_{XX} above with ϕ_{XX} after you"demean" the process.

Ergodicity

Motivation: So far we have based our description of SPs on a large number of sample functions (called an “ensemble”), i.e., we assume that we can reset the process as many times as we want, and then observe each sample function for a very very long time.

What if the ensemble is not available ? What if only one realization of the SP is all that is available?

Examples abound: earthquake excitations, stock market indices, annual rainfall at a location, dynamics of gas, etc.

From this one sample function we can not get the ensemble statistics of course: what we get are the “temporal statistics”. Can we say that a certain temporal average equals the corresponding ensemble average? If so, under what conditions?

If for a process the temporal and ensemble averages for a certain parameter are equal, the process is termed as ergodic in that parameter. As we will see, ergodicity requires stationarity. But all stationary processes need not be ergodic. In most cases, ergodicity needs to be assumed, instead of being proved. (Like statistical independence).

1.1.1 Mean value ergodicity

Consider the temporal mean value:

$$M_T = \overline{X(t)} = \frac{1}{2T}\int_{-T}^{T} X(t)dt$$

$$\text{(alternate form)} \quad = \frac{1}{T}\int_{0}^{T} X(t)\,dt \qquad \text{(0.}$$

where T is the length of the record, $x(t)$ is the single sample function.

The first question that naturally arises is, what do we do with this temporal average? Does

M_T approach $E\left(X(t)\right)$?

Obviously, this question can be entertained only if $E\left(X(t)\right)$ is independent of t. So we rephrase the question as: For a weakly stationary process, does $M_T \to \mu_X$ as $T \to \infty$?

If it does, then the process $X(t)$ is ergodic in the mean:

$$\lim_{T\to\infty} M_T = \mu_X \Rightarrow \text{mean value ergodic} \quad (0.2)$$

What are the conditions for this equality to hold?

Let us study the variance of the estimate M_T and its behaviour as the averaging window expands. Clearly, for a stationary process, M_T is an unbiased estimator since,

$$E(M_T) = \frac{1}{2T}\int_{-T}^{T} EX(t)dt = \frac{1}{2T}\mu_X 2T = \mu_X \qquad (0.$$

Its variance is:

$$\begin{aligned} \mathrm{var}(M_T) &= E(M_T^2) - E^2(M_T) = \frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T} EX(t_1)X(t_2)dt_1dt_2 \; - \mu_X^2 \\ &\vdots \qquad (0. \\ &= \frac{1}{2T}\int_{-2T}^{2T} G_{XX}(\tau)\left(1-\frac{|\tau|}{2T}\right)d\tau \end{aligned}$$

where K_{XX} is the ACVF of $X(t)$. M_T is a consistent estimator iff the right hand side of Eq (0.2) goes to zero as T grows large. Slutsky 's theorem (Papoulis p. 526) allows a more wider class of processes to be mean value ergodic:

$$X(t) \text{ is ergodic } \Leftrightarrow \lim_{T\to\infty} \frac{1}{T}\int_0^T G_{XX}(\tau)\,d\tau = 0 \quad (0.3)$$

Examples

Examples:

$X(t) = C$ (random horizontal line)

$\mu_X = \mu_C$

But $M_T \neq \mu_C$ for almost all sample paths of $X(t)$

Hence $X(t)$ is not mean ergodic

(0.

$$G_{XX}(\tau) = \sigma^2 \exp(-|\tau|/c)$$

$$\text{var}(M_T) = \frac{\sigma^2}{T}\int_0^{2T} \exp(-\tau/c)(1-\frac{\tau}{2T})d\tau$$

$$= \frac{\sigma^2 c}{T}\left(1 - \frac{1-\exp(-2T/c)}{2T/c}\right)$$

$$\to 0 \text{ if } T >> c$$

(0.

$X(t) = A\cos wt + B\sin wt + c$

A, B are zero mean, unit variance, independent RVs

$E(X(t)) = c$

$\text{var}(X(t)) = 1$

(0.

$G_{XX}(\tau) = \cos w\tau$

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T G_{XX}(\tau)\,d\tau = \lim_{T\to\infty} \frac{1}{T}\sin wt = 0$$

$X(t)$ is mean ergodic

$X(t) = c + \xi(t)$

$M_T = c$

$$\text{var}(M_T) = \frac{1}{2T}\int_{-2T}^{2T} \delta(t)(1-\frac{|\tau|}{2T})d\tau = \frac{1}{2T} \to 0 \qquad (0.4)$$

$X(t)$ is mean ergodic

Markov processes

Markov assumption

1.1.1 The Markov property

Let a stochastic process $X(t)$ be observed at arbitrary but ordered times $t_1 < t_2 < ... < t_{k-1} < t_k < t_{k+1} < ... < t_n$. If the conditional probability of the state at time t_{k+1} given the k observations at $t_1, t_2, ..., t_k$ equals the conditional probability given the most recent of these k observations,

$$P[X(t_{k+1}) = x_{k+1} \mid X(t_k) = x_k, X(t_{k-1}) = x_{k-1}, ..., X(t_2) = x_2, X(t_1) = x_1] = P[X(t_{k+1}) = x_{k+1} \mid X(t_k) = x_k] \quad (0.$$

then $X(t)$ is a Markov process.

Markov processes

A general continuous state continuous time Markov process can be described by the differential form of the Chapman Kolmogorov equation:

$$\frac{\partial}{\partial t}p(z,t\,|\,y,t') = -\frac{\partial}{\partial z}A(z,t)\,p(z,t\,|\,y,t') + \frac{1}{2}\frac{\partial^2}{\partial z^2}B(z,t)\,p(z,t\,|\,y,t') + \int\left[W(z\,|\,x,t)\,p(x,t\,|\,y,t') - W(x\,|\,z,t)\,p(z,t\,|\,y,t')\right]dx \qquad (0.$$

The initial condition is $p(z,t\,|\,y,t) = \delta(y-z)$. By "integrating out" the conditioning event $X(t') = y$, Eq (0.1) can be given a more pleasing appearance:

$$\frac{\partial}{\partial t}p(z,t) = -\frac{\partial}{\partial z}A(z,t)\,p(z,t) + \frac{1}{2}\frac{\partial^2}{\partial z^2}B(z,t)\,p(z,t) + \int\left[W(z\,|\,x,t)\,p(x,t) - W(x\,|\,z,t)\,p(z,t)\right]dx \qquad (0.$$

with an initial condition that is more well-behaved: $P(X(t_0) = x_0) = p(x_0,t_0)$.

The first term is the *drift* term:

$$A(z,t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|x-z|<\varepsilon} (x-z) p(x, t+\Delta t \mid z, t) dx$$
$$= \lim_{\Delta t \to 0} \frac{1}{\Delta t} E[X(t+\Delta t) - X(t) \mid X(t) = z] \quad (0.$$

The second term is the *diffusion* term

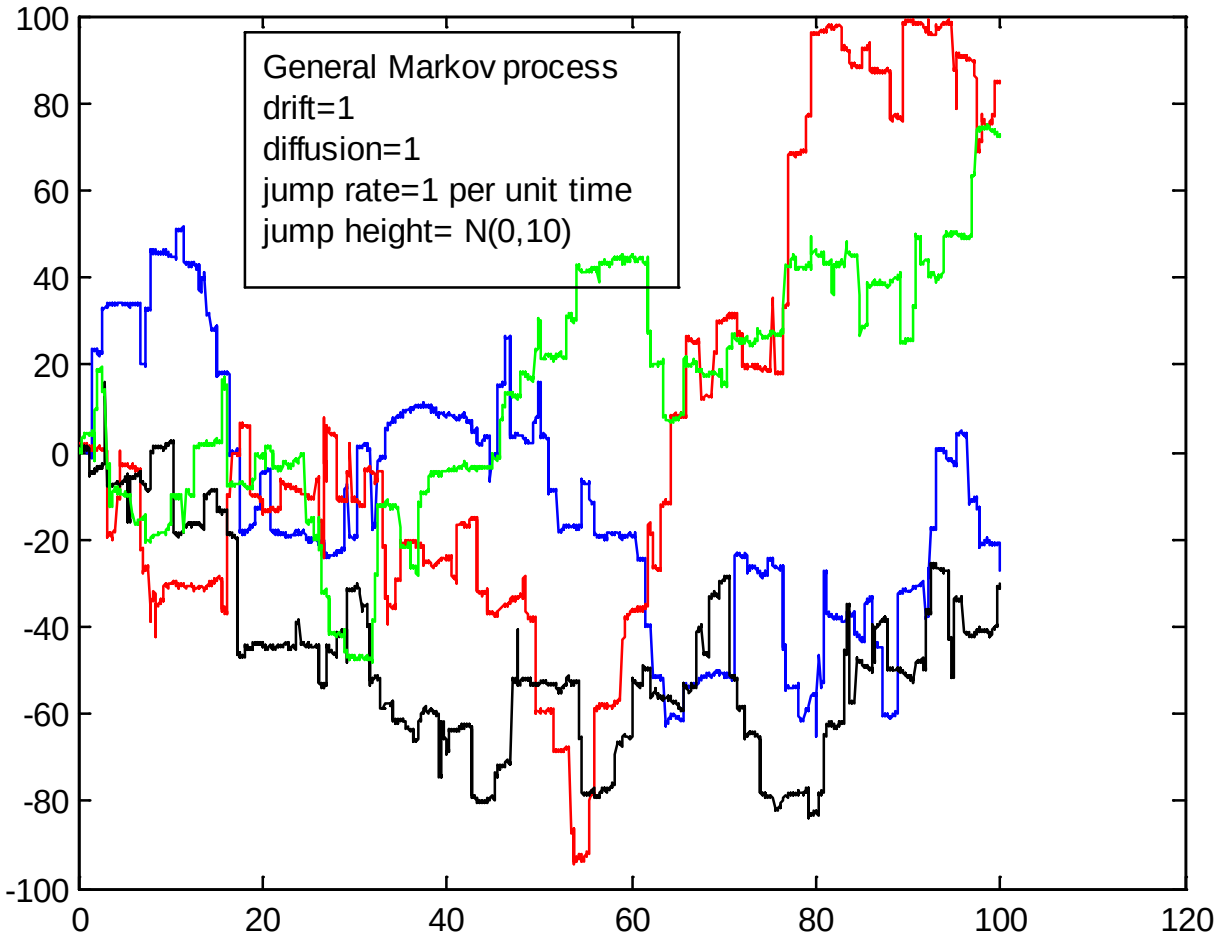
$$B(z,t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|x-z|<\varepsilon} (x-z)^2 p(x, t+\Delta t \mid z, t) dx$$
$$= \lim_{\Delta t \to 0} \frac{1}{\Delta t} E[\{X(t+\Delta t) - X(t)\}^2 \mid X(t) = z] \quad (0.$$

The third term is the *jump* term. The jump rate is given by

$$W(x \mid z,t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} p(x, t+\Delta t \mid z, t) \quad (0.$$

It is assumed that A and B converge uniformly in z, ε, t; and W converges uniformly in x, z and t for all $\varepsilon > 0$ such that $|x - z| \geq 0$.

Example: drift diffusion and jump



Fokker Planck equation

Recall that the jump term is non -existent in a diffusion process. The Chapman Kolmogorov differential equation thus becomes

$$\frac{\partial}{\partial t} p(z,t) = -\frac{\partial}{\partial z} A(z,t)\, p(z,t) + \frac{1}{2}\frac{\partial^2}{\partial z^2} B(z,t)\, p(z,t) \qquad (0.$$

which is known as the Fokker Planck equation with initial condition $P(X(t_0) = x) = p(x,t_0)$.

The equivalent Ito stochastic differential equation is

$$dz = A(z,t)dt + \sqrt{B(z,t)}dW(t) \qquad (0.2)$$

Wiener process

Examples:

Wiener process. A=0, B=1.

FPE becomes,

$$\frac{\partial}{\partial t}p(z,t)=\frac{1}{2}\frac{\partial^2}{\partial z^2}p(z,t) \qquad (0.$$

Heat equation with conductivity ½. Initial condition: $p(z,0)=\delta(z)$.

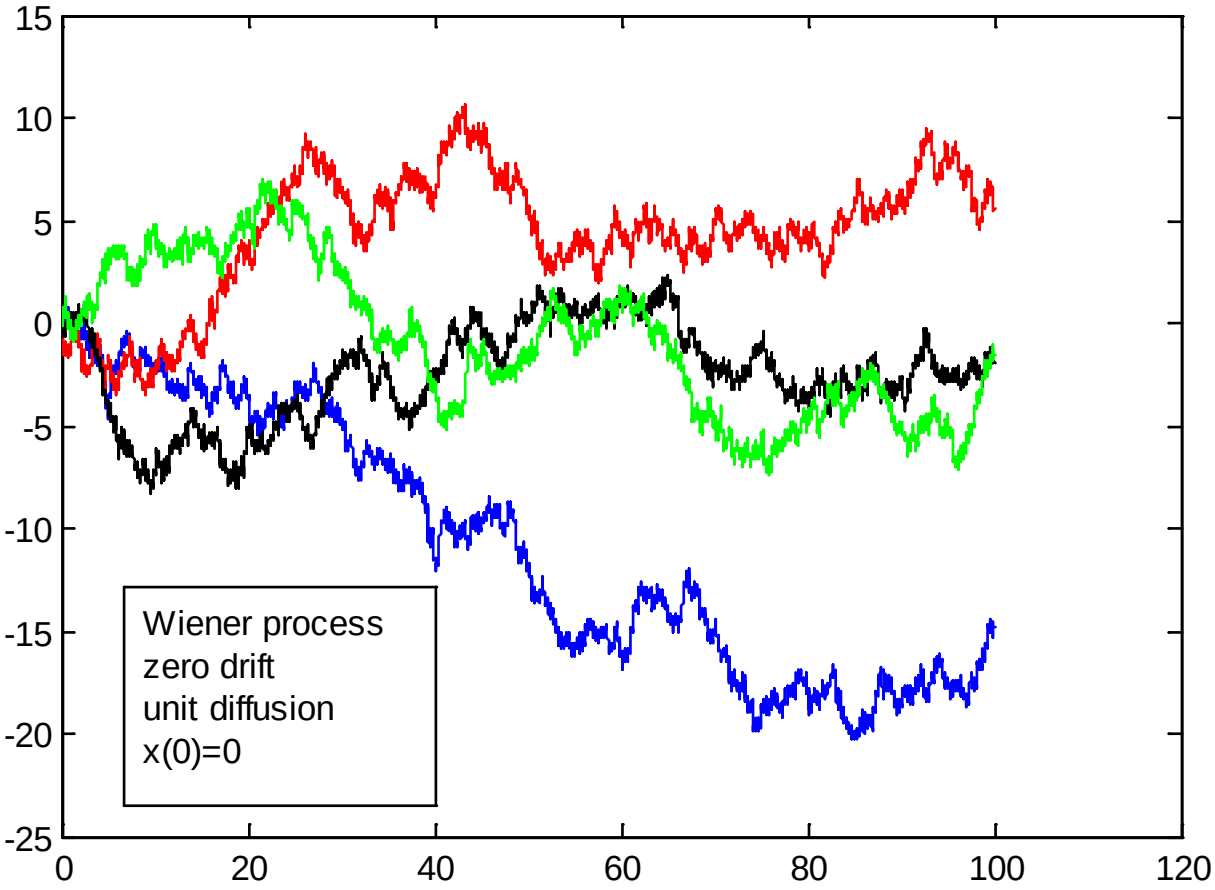
Solution is:

$$p(z,t)=\frac{1}{\sqrt{4\pi\alpha t}}\int_{-\infty}^{\infty}\exp[-(z-y)^2/(4\pi\alpha t)]\delta(y)dy$$

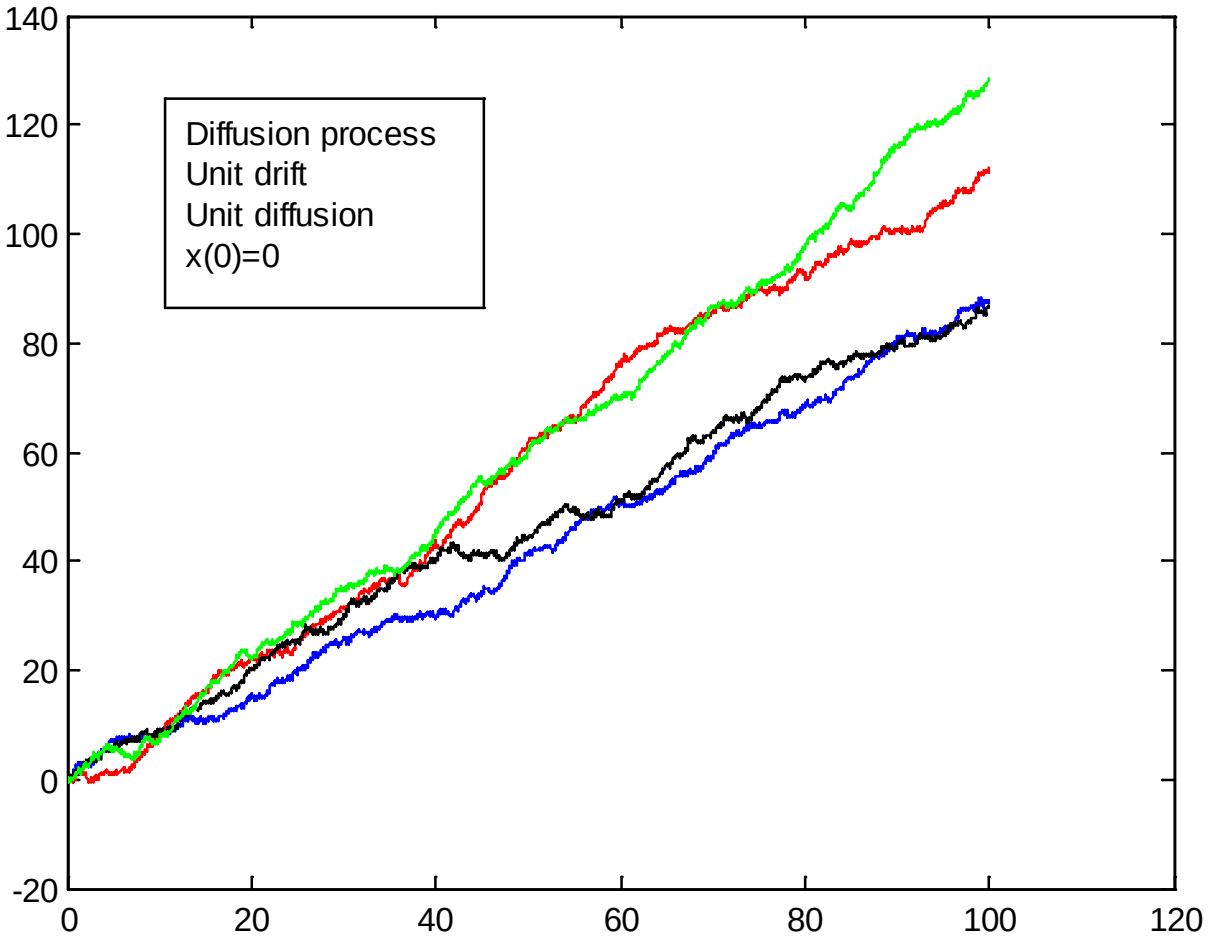
$$=\frac{1}{\sqrt{2\pi t}}\exp[-(z)^2/(2\pi t)]$$

$$\therefore\ Z(t)\sim N(0,\sqrt{t})$$

$$K_{ZZ}(t,s)=\min(s,t)$$



Example


Diffusion process
Unit drift
Unit diffusion
x(0)=0
140
120
100
80
60
40
20
0
-20
0
20
40
60
80
100
120

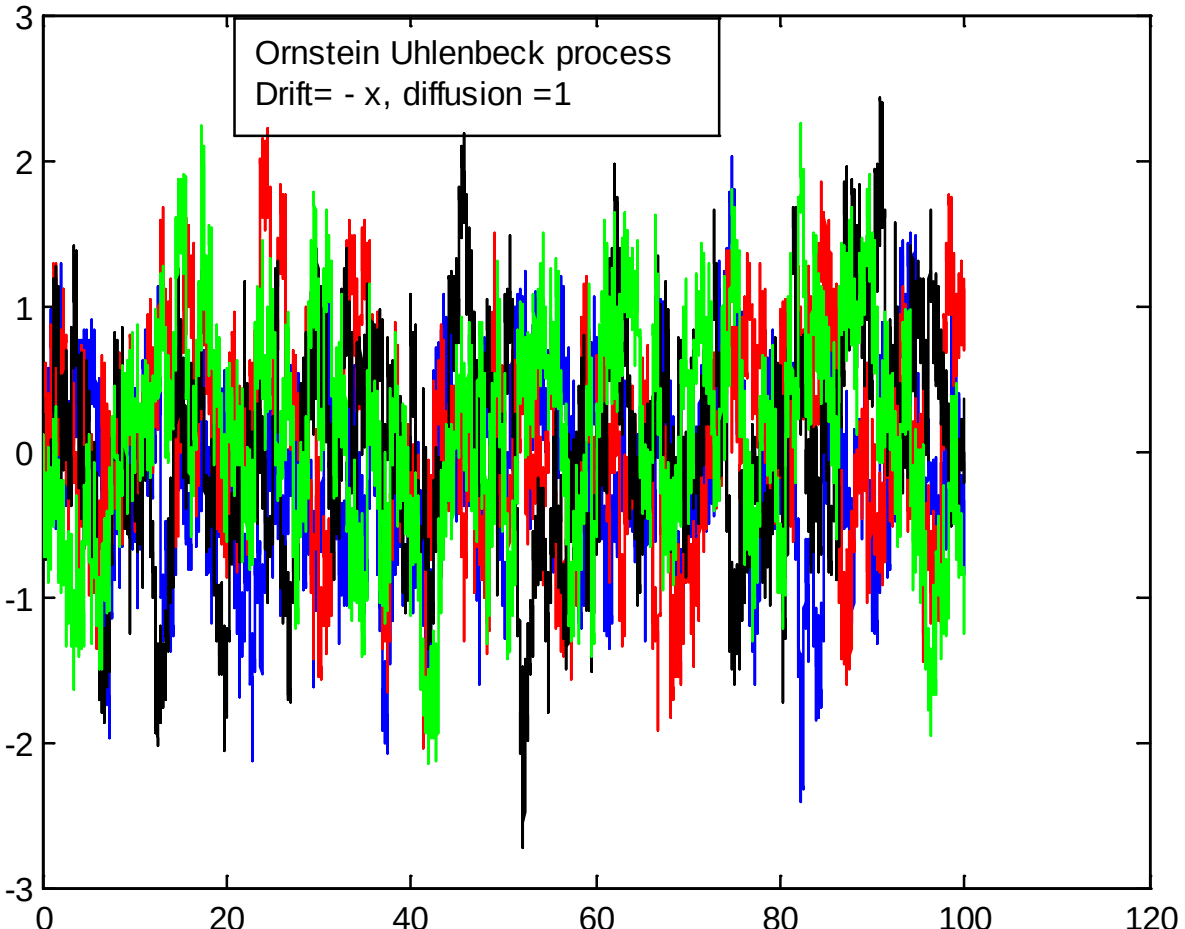
Example

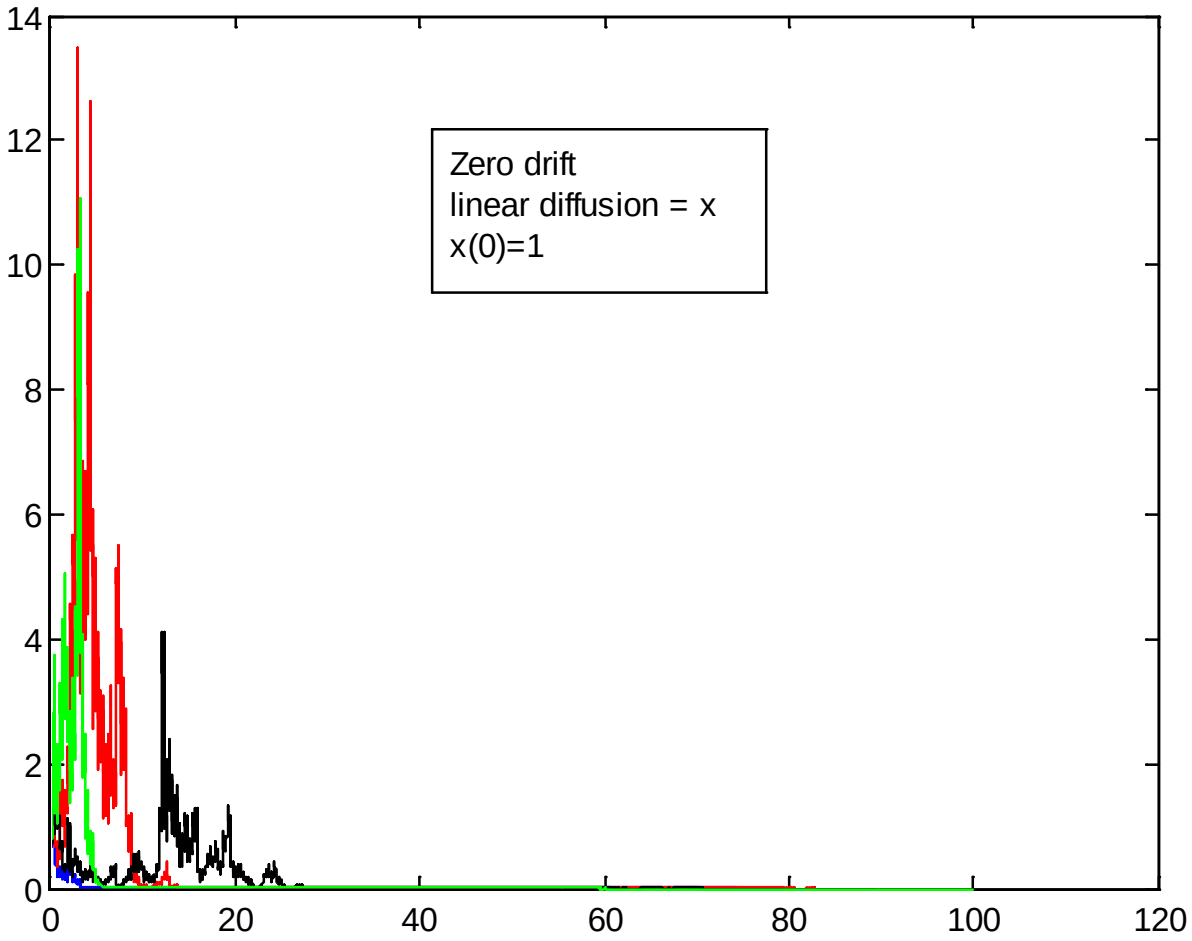
Ornstein-Uhlenbeck process. A= - cz, B=1.

Solution:

$$Z(t) \sim N(z_0 e^{-ct}, \frac{1}{2c}[1-e^{-2ct}]) \qquad (0.1)$$

$$K_{ZZ}(s,t) \to \frac{1}{2c} e^{-c|t-s|}$$




Zero drift
linear diffusion = x
x(0)=1
0
2
4
6
8
10
12
14
0
20
40
60
80
100
120

Deterministic process

1.1.1.1 Deterministic process

If jump and diffusion are zero, the process is deterministic. The form of the FPE then becomes that of Liouville's equation:

$$\frac{\partial}{\partial t} p(\underline{z},t) = -\sum_i \frac{\partial}{\partial z_i} A_i(\underline{z},t)\, p(\underline{z},t) \qquad (0.$$

that is,

$$\frac{dz_i}{dt} = A_i(\underline{z},t) \quad (0.2)$$

Detailed balance

1.1.1.1 Detailed balance

At time t, phase point is $\Gamma(x,v;t)$

At time t+τ, it evolves to $\Gamma(x',v';t+\tau)$

The joint probability of these two events is: $p(x',v',t+\tau;x,v,t)$

Reset: at time t, phase point is $\Gamma(x',-v';t)$

Does phase point at time t+τ evolve to $\Gamma(x,-v;t+\tau)$?

The joint probability of these two events is: $p(x,-v,t+\tau;x',-v',t)$

Are these two probabilities equal? If so, we have detailed balance.

For a stationary Markov process, the two probabilities are respectively:

$p(x',v',\tau;x,v,0)\,p_s(x,v)$ and $p(x,-v,\tau;x',-v',0)\,p_s(x',-v')$

Consequence of detailed balance

1.1.1.1 Consequence of detailed balance

Define $\varepsilon_i = \pm 1$ for replacing z_i with $\varepsilon_i z_i$ depending on whether z_i is odd or even.

Detailed balance then can be written as:

$$p(\underline{z}, \tau \mid \underline{z}', 0)\, p_s(\underline{z}') = p(\underline{\varepsilon z}', \tau \mid \underline{\varepsilon z}, 0)\, p_s(\underline{\varepsilon z})$$

Also,

$$p_s(\underline{\varepsilon z}) = p_s(\underline{z}) \qquad (0.1)$$

Then for FPE,

$$\varepsilon_i A_i(\underline{\varepsilon z})\, p_s(\underline{z}) = -A_i(\underline{z})\, p_s(\underline{z}) + \sum \frac{\partial}{\partial z_j} B_{ij}(\underline{z})\, p_s(\underline{z})$$

$$\varepsilon_i \varepsilon_j B_{ij}(\underline{\varepsilon z}) = B_{ij}(\underline{z})$$

1.1.1.1 Kramer's equation for Brownian motion

The SDE is:

$$\frac{dx}{dt} = v$$
$$m\frac{dv}{dt} = -\Phi'(x) - \eta v + \sqrt{2\eta k_B T}\,\xi(t) \qquad (0.$$

The corresponding FPE terms are:

$$A = \begin{bmatrix} v \\ -\frac{1}{m}[\Phi'(x) + \eta v] \end{bmatrix},\ B = \begin{bmatrix} 0 & 0 \\ 0 & 2\eta k_B T \end{bmatrix} \qquad (0.$$

Canonical distribution

Using the first result of detailed balance, we get:

$$\underline{\varepsilon}\,\underline{A}(x,-v)\,p_s(x,v) = -A(x,v)\,p_s(x,v) + \begin{bmatrix} 0 \\ 2\eta \dfrac{\partial p_s}{\partial v} \end{bmatrix}$$

Substituting,

$$\begin{bmatrix} -v \\ \Phi'(x) - \eta v \end{bmatrix} p_s = \begin{bmatrix} -v \\ \Phi'(x) + \eta v \end{bmatrix} p_s + \begin{bmatrix} 0 \\ 2\eta \dfrac{\partial p_s}{\partial v} \end{bmatrix}$$

The second yields, (0.

$$-v p_s = \frac{\partial p_s}{\partial v}$$

that is,

$$p_s = \exp[-\frac{v^2}{2}]\, f(x)$$

Substituting this stationary Markov probability into the original FPE, we get:

$$0 = -v\frac{\partial f}{\partial x} - \Phi'(x) v f$$

which gives,

$$f(x) = C\exp[-\Phi(x)] \qquad (0.2)$$

Thus,

$$p_s(x,v) = C\exp\left[\frac{-\Phi(x)}{k_B T} - \frac{m v^2}{2 k_B T}\right]$$

Thank you