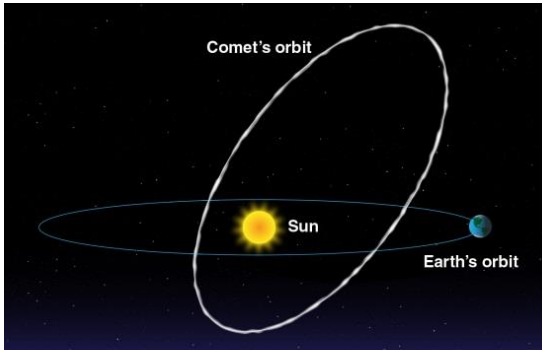
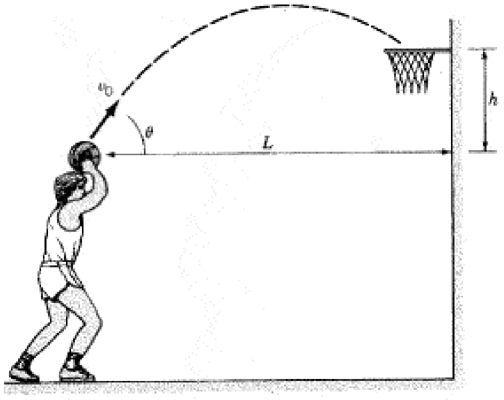
Probability theory

What predictions can we make about systems based on incomplete information?

Deterministic systems: *Future behavior of the system* (*events*) *can be predicted with a required accuracy;*

• Classical mechanical systems with known initial conditions





Stochastic systems: *Future behavior of the system* (*events*) *cannot be predicted with certainty;*

• Mechanical systems with incomplete knowledge of initial conditions (Δx, Δv) or unknown external forces

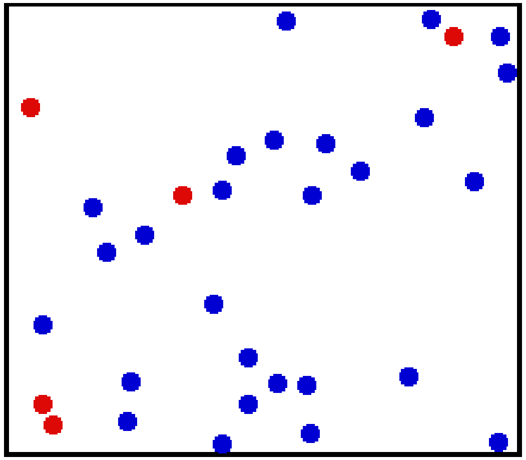
Probability distributions *are assigned to stochastic systems to represent as much information about the system as possible.*

Probability theory for molecular systems

Molecular systems are deterministic so why do we need probability distributions to describe them?! Two cases apply:

1) The system is too large or complex to determine and/or keep track of all mechanical variables and their time dependence

2) In molecular simulations there is the opposite problem of too much information!



Wikipedia: Kinetic theory of gases

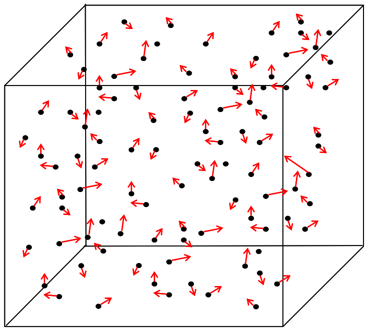
- To make predictions about the bulk behavior of the system, are the details of the motion of every single atom needed?
- Are there average values of mechanical properties which capture the macroscopic behavior of the system?
- Can we determine these macroscopic averages without knowing all the microscopic details?

Probability distributions for molecular systems

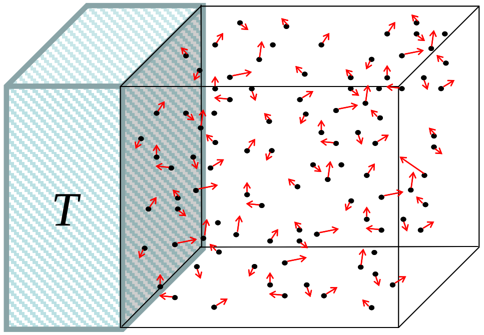
1) What form do probability distributions take for molecular systems? How do they depend on positions, velocities, energies, … of the molecules?

2) Is there a unique probability distribution for molecular variables for each system?

3) How is the probability distribution affected by interactions of the system with the environment?



Isolated system



System with thermostat

Statistical mechanics gives us a methodology for constructing probability distributions for the mechanical properties, without having to solve for all the microscopic mechanical variables.

- What are the properties of probability distributions?
- How do probability distribution behave for large numbers of variables?

Probability distributions for systems with discrete variables

- The outcome of an event cannot be predicted with certainty
- Total number of possible outcomes of the measurement are finite (ν)

The probability distribution represents our knowledge of the system:

$$\begin{pmatrix} \varepsilon(1) & \varepsilon(2) & \varepsilon(3) & \cdots & \varepsilon(\nu) \\ P(1) & P(2) & P(3) & \cdots & P(\nu) \end{pmatrix}$$

← Possible events ("event space") are assigned numbers

← A probability is associated with each event

Conditions a probability distribution must satisfy:

1) For all events *i* assigned probabilities are positive:

$$P(i) \geq 0$$

2) The sum of the probabilities add up to 1

$$\sum_{i=1}^{\nu} P(i) = 1$$

Examples of probability distributions for systems with discrete variables

• Physical events ε_i are assigned probabilities $P(i)$ which reflect their nature as much as possible. The events are also assigned numbers.

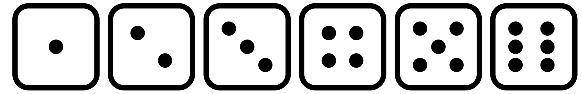
Flipping coins

ε_i	0	1
$P(i)$	$\frac{1}{2}$	$\frac{1}{2}$

Poker

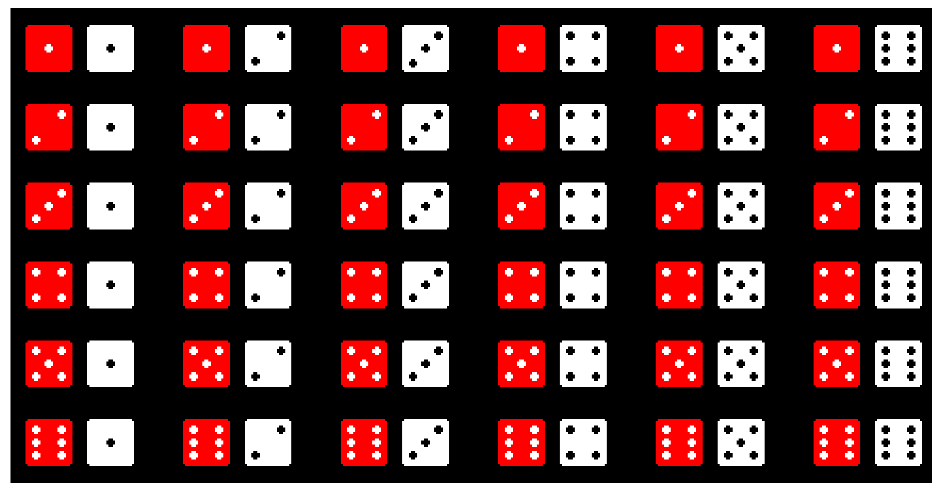
Role of one die

ε_i



$P(i)$ $\frac{1}{6}$ $\frac{1}{6}$ $\frac{1}{6}$ $\frac{1}{6}$ $\frac{1}{6}$ $\frac{1}{6}$

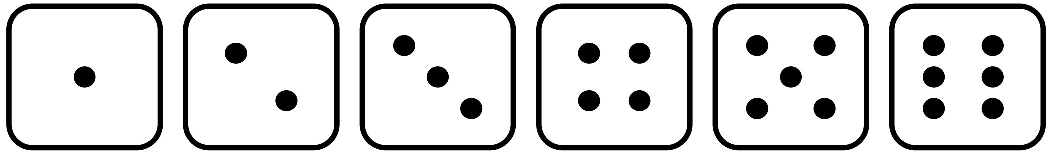
Rolling two dice



E_i	1	2	3	…	36
$P_2(i)$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$		$\frac{1}{36}$

Characterizing a probability distribution with discrete variables

Event ε_i



Probability $P(i)$ $\frac{1}{6}$ $\frac{1}{6}$ $\frac{1}{6}$ $\frac{1}{6}$ $\frac{1}{6}$ $\frac{1}{6}$

How can we characterize the distribution?

"Measures" of the distribution:

1) *Average* or *expectation value* for a distribution

$$\langle \varepsilon \rangle = \bar{\varepsilon} = \sum_{i=1}^{\nu} \varepsilon_i P(i)$$

For the case of 1 die:

$$\langle \varepsilon \rangle = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$$

Note: The average of the distribution does not have to correspond to an event

Characterizing probability distributions

2) *"Moments"* of a distribution (or the related "cumulants")

- 2nd moment of the distribution for the throw of one die $\left\langle \varepsilon^2 \right\rangle = \sum_{i=1}^{\nu} \varepsilon_i^2 P(i)$

$$\left\langle \varepsilon^2 \right\rangle = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6}$$
$$= 15.167$$

Note: $\sqrt{\left\langle \varepsilon^2 \right\rangle} = 3.89 \neq \left\langle \varepsilon \right\rangle$

- 2nd central moment (*variance*) $\left\langle \left(\varepsilon_i - \left\langle \varepsilon \right\rangle \right)^2 \right\rangle = \sum_{i=1}^{\nu} \left(\varepsilon_i - \left\langle \varepsilon \right\rangle \right)^2 P(i)$

$$\left\langle \left(\varepsilon - \left\langle \varepsilon \right\rangle \right)^2 \right\rangle = (1-3.5)^2 \cdot \frac{1}{6} + (2-3.5)^2 \cdot \frac{1}{6} + (3-3.5)^2 \cdot \frac{1}{6} + (4-3.5)^2 \cdot \frac{1}{6} + (5-3.5)^2 \cdot \frac{1}{6} + (6-3.5)^2 \cdot \frac{1}{6}$$
$$= 2.9167$$

The standard deviation characterizes the *width* of a distribution

$$\sqrt{\left\langle \left(\varepsilon - \left\langle \varepsilon \right\rangle \right)^2 \right\rangle} = 1.7078 = \sigma$$

Characterizing probability distributions

Higher order moments of a distribution are required to fully characterize a distribution

- The *m^{th} moment* of a distribution

$$\left\langle \varepsilon^m \right\rangle = \sum_{i=1}^{\nu} \varepsilon_i^m P(i)$$

- The *m^{th} central moment* of a distribution

$$\left\langle \left(\varepsilon_i - \left\langle \varepsilon \right\rangle \right)^m \right\rangle = \sum_{i=1}^{\nu} \left(\varepsilon_i - \left\langle \varepsilon \right\rangle \right)^m P(i)$$

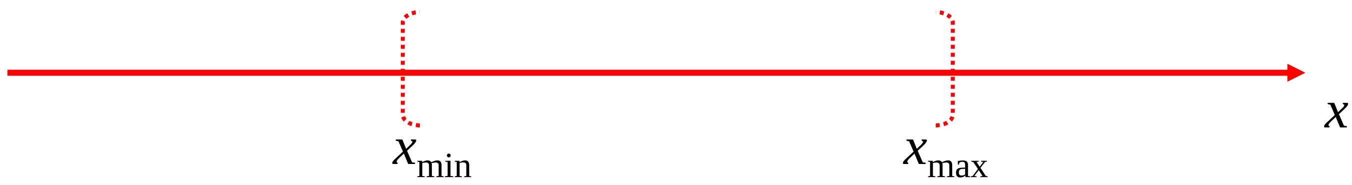
- The average of any mathematical function of ε_i for the distribution:

$$\left\langle f(\varepsilon) \right\rangle = \sum_{i=1}^{\nu} f(\varepsilon_i) P(i)$$

Probability distributions with continuous variables

• Stochastic variables (events) take continuous values within a range $[x_{min}, x_{max}]$

An infinite number of x values are possible in the range



Common ranges for variables: $0 \leftrightarrow 1$

$0 \leftrightarrow +\infty$

$-\infty \leftrightarrow +\infty$

• There are an infinite number of points on the real number axis, the probability of getting an exact value x from a measurement is mathematically 0.

• Probability of observing the event between x and x+dx is given as $P(x)dx$ where $P(x)$ is the probability distribution for a continuous variable,

Properties of the probability distribution $P(x)$ for a continuous variable:

1) All probabilities are positive

For all x : $P(x) \geq 0$

2) The probability is normalized

$$\int_{x_{min}}^{x_{max}} P(x)dx = 1$$

Characterizing continuous probability distributions

- *Average* or *expectation value* for a distribution

$$\langle x \rangle = \bar{x} = \int_{x_{\min}}^{x_{\max}} x P(x) dx$$

- The *2nd moment* of the distribution is

$$\langle x^2 \rangle = \int_{x_{\min}}^{x_{\max}} x^2 P(x) dx$$

- The *2ne central moment* or *variance* of the distribution

$$\left\langle \left(x - \langle x \rangle \right)^2 \right\rangle = \int_{x_{\min}}^{x_{\max}} \left(x - \langle x \rangle \right)^2 P(x) dx$$

- The *standard deviation* of the distribution

$$\sqrt{\left\langle \left(x - \langle x \rangle \right)^2 \right\rangle} = \sigma$$

Characterizing continuous probability distributions

- The m^{th} *moment* of the distribution

$$\left\langle x^m \right\rangle = \int_{x_{\min}}^{x_{\max}} x^m P(x)dx$$

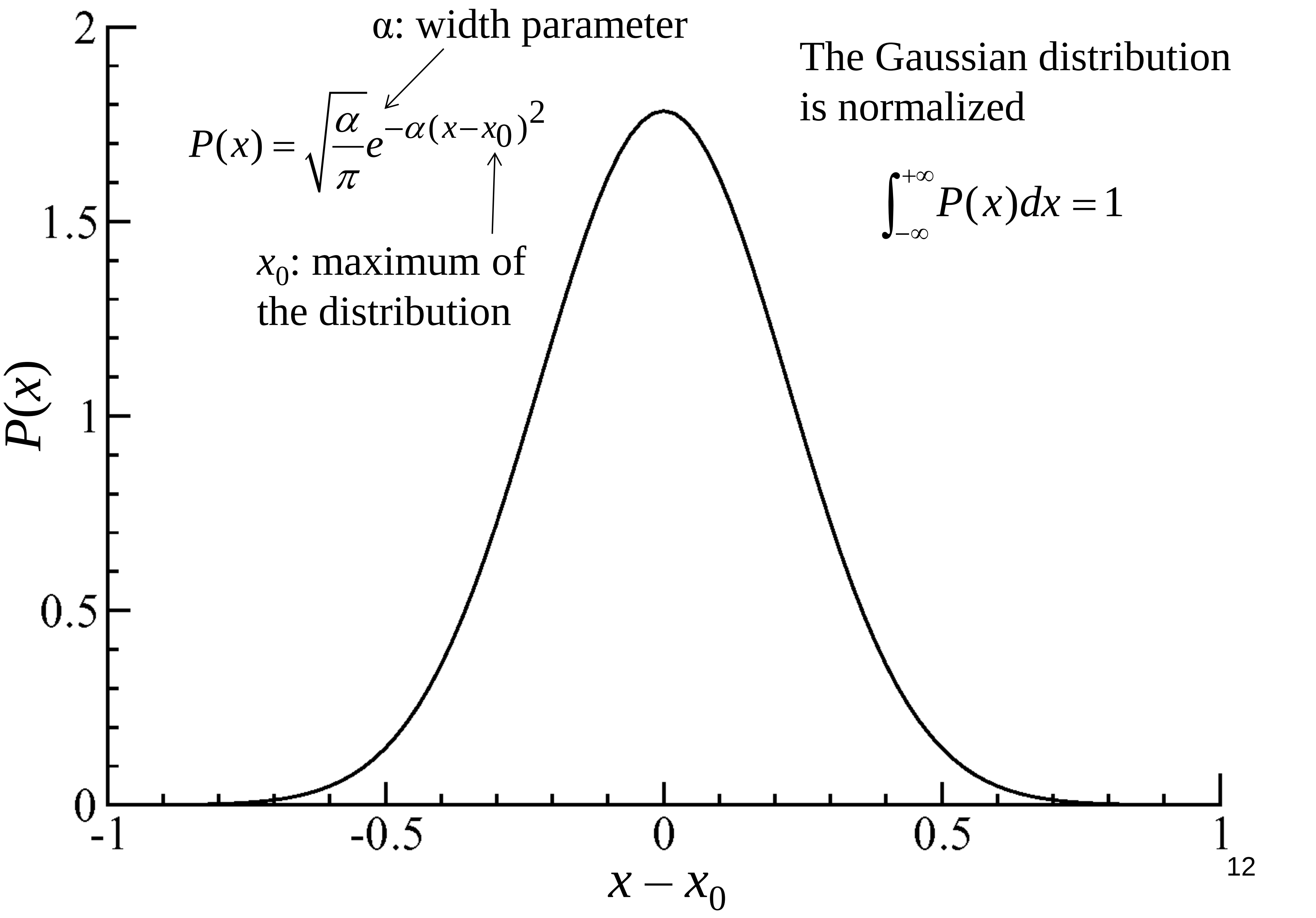
- The m^{th} *central moment* of the distribution

$$\left\langle \left(x - \langle x \rangle \right)^m \right\rangle = \int_{x_{\min}}^{x_{\max}} \left(x - \langle x \rangle \right)^m P(x)dx$$

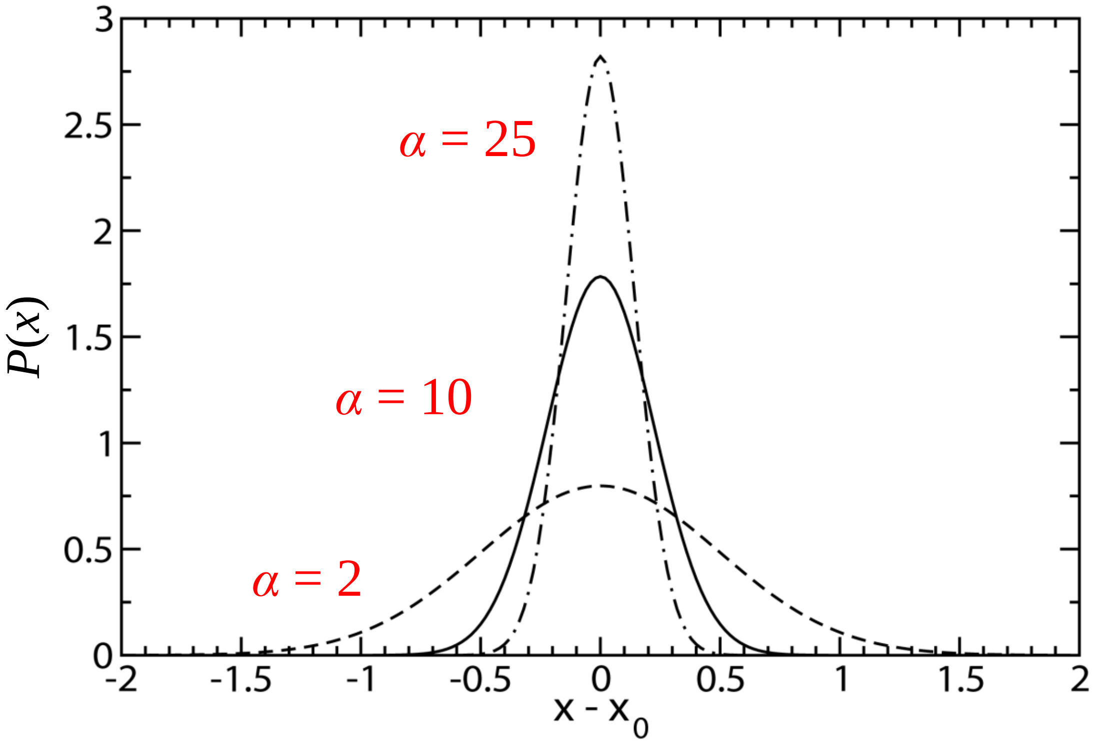
- The average of any function $f(x)$ of x can be calculated for the distribution

$$\left\langle f(x) \right\rangle = \int_{x_{\min}}^{x_{\max}} f(x)P(x)dx$$

The Gaussian distribution: a widely used continuous distribution function



The Gaussian ("normal") distribution function



$$P(x) = \sqrt{\frac{\alpha}{\pi}} e^{-\alpha(x-x_0)^2}$$

- Considered the "natural" or "normal" distribution function when no other information is known about the distribution of the variables

Average of x in Gaussian distribution:

$$\langle x \rangle = \int_{-\infty}^{+\infty} x P(x) dx = \sqrt{\alpha/\pi} \int_{-\infty}^{+\infty} x e^{-\alpha(x-x_0)^2} dx = x_0$$

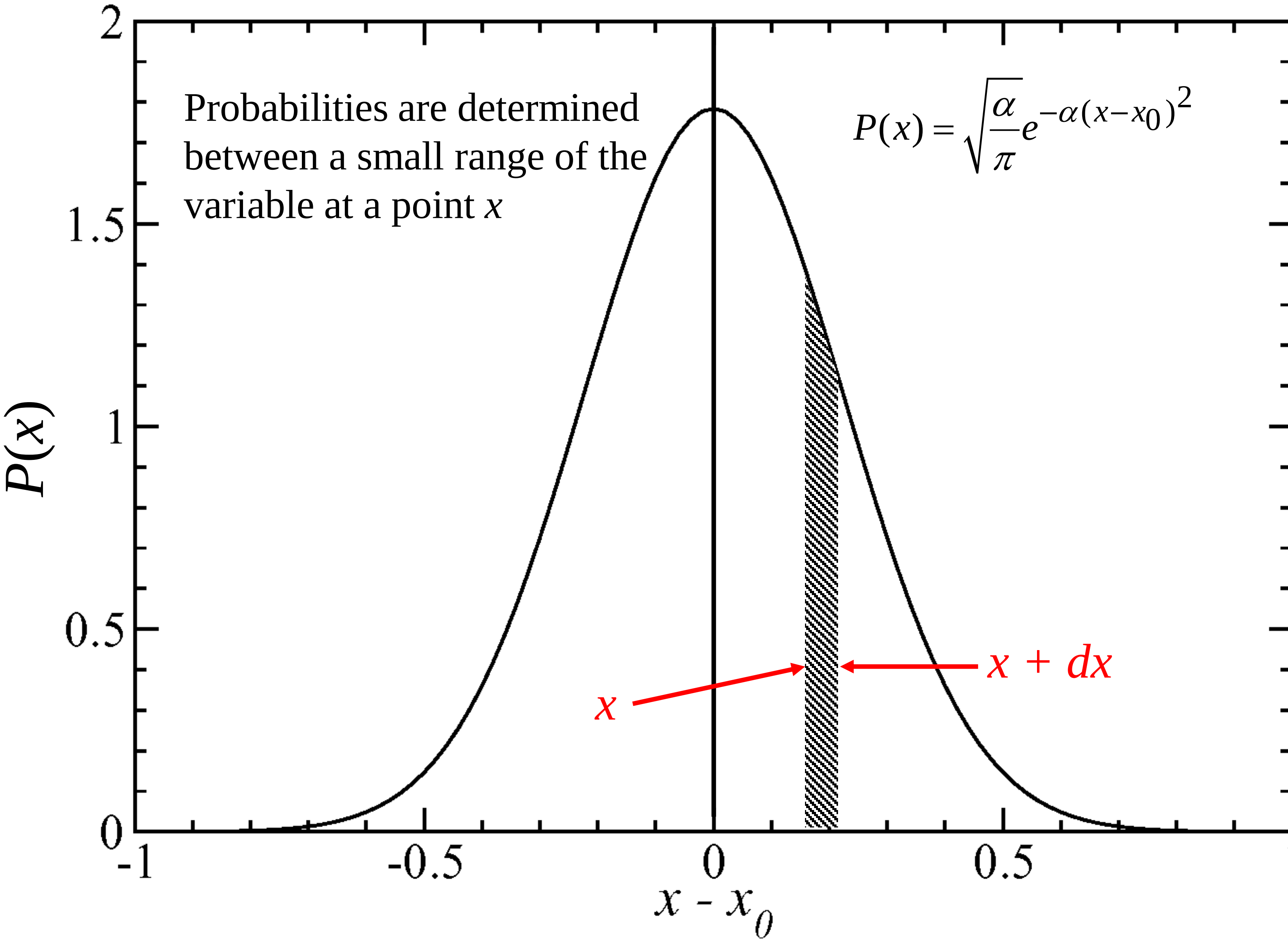
Variance of x in a Gaussian distribution:

$$\left\langle \left(x - \langle x \rangle\right)^2 \right\rangle = \frac{1}{2\alpha} = \sigma^2$$

Note: the three distributions above have the same average, but not the same variance

See Appendices of Chapter 5 for integrals

The Gaussian distribution



$P(x)dx$ = Probability of observing the variable between x and $x + dx$

Distributions functions of many independent variables

For a system with N variables $x_1, x_2, \ldots, x_N$ we can have **collective variables** X, which are a property of the entire system

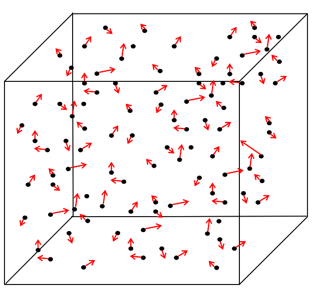
$P_N(X)$ is the probability distribution of collective variable X.

A simple case of a collective variable is the sum of the individual variables:

$X = x_1 + x_2 + \ldots + x_N$

- X represents a "macrostate" or collective property of the system
- The set of individual variables $\{x_1, x_2, \ldots, x_N\}$ associated with a particular X are called the "microstate"

-This type of problem arises in molecular systems: What is the probability of the ideal gas system has an energy E



$E_N = \varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_N$

What is $P_N(E)$?

Probability distributions with multiple independent stochastic variables

Independent stochastic variables $x_1, x_2, \ldots x_N$ have distribution functions

$$P_1(x_1), P_1(x_2), \ldots, P_1(x_N)$$

The variables $x_1, x_2, \ldots, x_N$ are similar in nature and have the same mathematical form for their probability distribution, $P_1(x_i)$

The N-variable probability distribution is

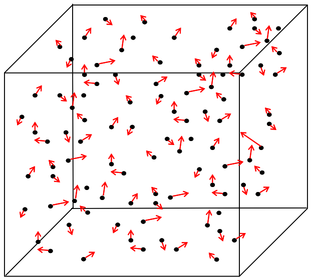
$$P_N(x_1,x_2,\ldots x_N) = P_1(x_1)\cdot P_1(x_2)\cdot\ldots\cdot P_1(x_N)$$

The probability distribution for $X = x_1 + x_2 + \ldots + x_N$ is

$$P_N(X) = P_N(x_1, x_2, \cdots, x_N) = P_1(x_1)P_1(x_2)\cdots P_1(x_N)$$

Note that different combinations of $x_1, x_2, \ldots, x_N$ can give the same X

- *Degeneracy*

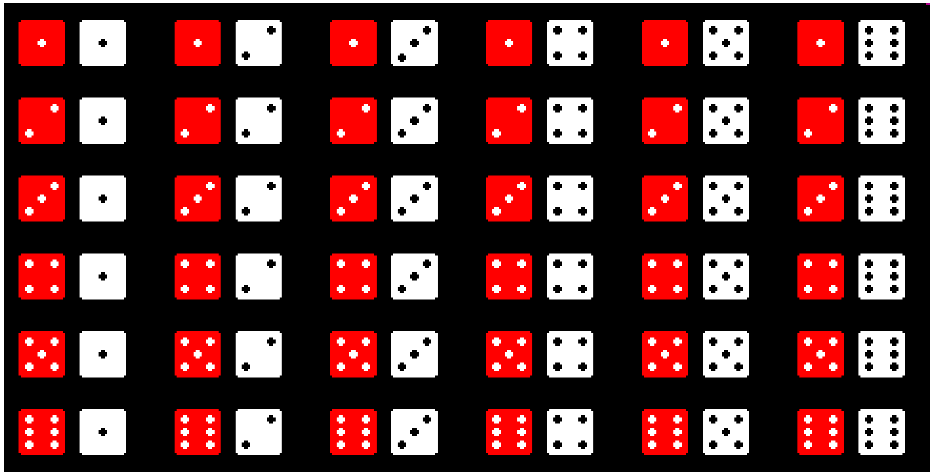


$$E_N = \varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_N$$

$$P_N(E) = \sum{}' P_1(\varepsilon_1)\, P_1(\varepsilon_2)\ldots P_1(\varepsilon_N)$$

Two-variable distribution with independent variables

Role of two dice



Observed outcome variable is a sum:

$E = \varepsilon_{\mathrm{I}} + \varepsilon_{\mathrm{II}}$

Observed probability is a product:

$P_2(E) = \sum' P_1(\varepsilon_{\mathrm{I}}) P_1(\varepsilon_{\mathrm{II}})$

Sum is over all ε_{I} and $\varepsilon_{\mathrm{II}}$ which give $E = \varepsilon_{\mathrm{I}} + \varepsilon_{\mathrm{II}}$

Possible individual events, $(\varepsilon_{\mathrm{I}}, \varepsilon_{\mathrm{II}})$ (microstates)	Value of $E=\varepsilon_{\mathrm{I}}+\varepsilon_{\mathrm{II}}$ (macrostate)	Degeneracy of macrostate, $\Omega(E)$	$P_2(E)$
(1,1)	2	1	1/36
(1,2) (2,1)	3	2	2/36
(1,3) (2,2) (3,1)	4	3	3/36
(1,4) (2,3) (3,2) (4,1)	5	4	4/36
(1,5) (2,4) (3,3) (4,2) (5,1)	6	5	5/36
(1,6) (2,5) (3,4) (4,3) (5,2) (6,1)	7	6	6/36
(2,6) (3,5) (4,4) (5,3) (6,2)	8	5	5/36
(3,6) (4,5) (5,4) (6,3)	9	4	4/36
(4,6) (5,5) (6,4)	10	3	3/36
(5,6) (6,5)	11	2	2/36
(6,6)	12	1	1/36

Convolution of probabilities: Multi-variable distributions from one-variable distributions

How do we determine the probability of a particular value of X from the role of the N separate dice?

For two dice, what is the probability of rolling a 4?

$$P_2(4) = P_1(1)P_1(3) + P_1(2)P_1(2) + P_1(3)P_1(1) = \frac{3}{36}$$

All cases for which $x_1 + x_2 = 4$?

Formalizing this expression for rolling any value:

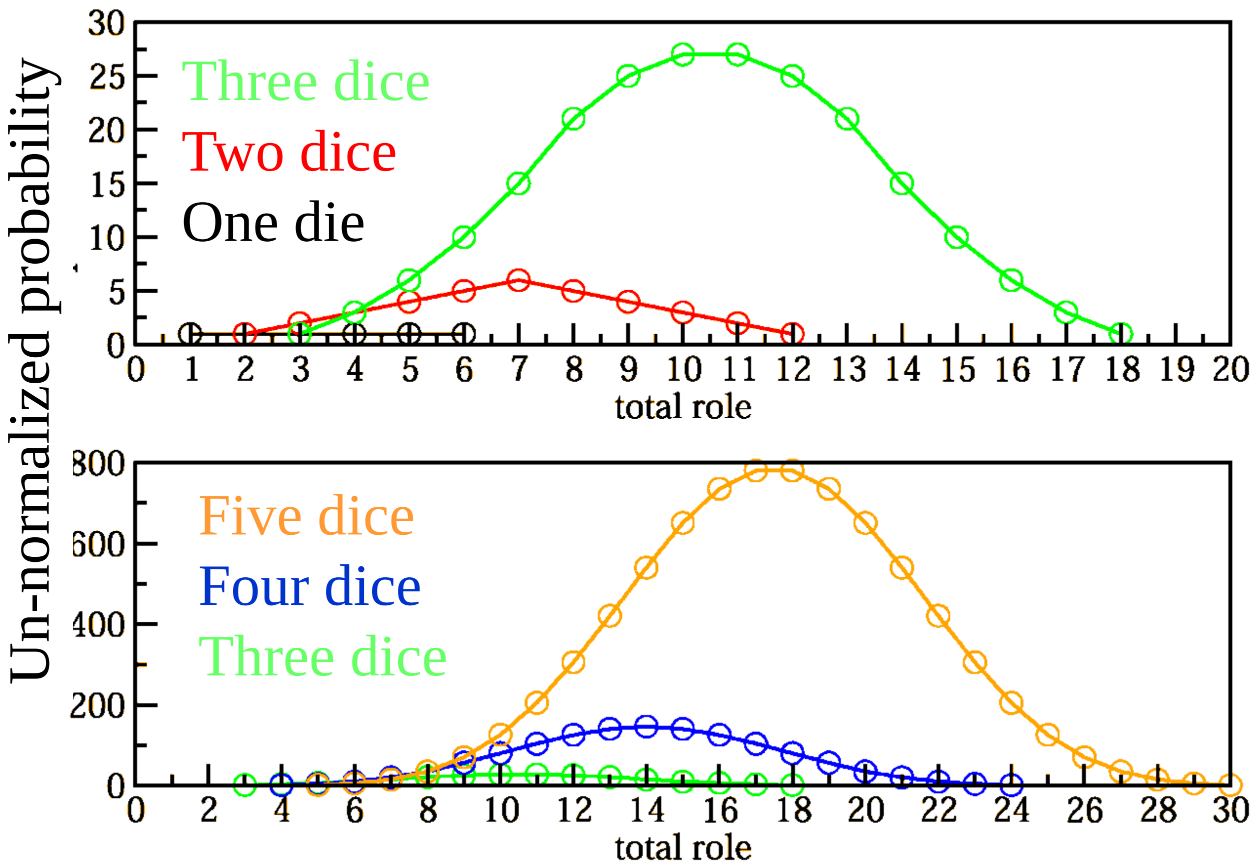
$$P_2(E) = \sum_{\varepsilon_I, \varepsilon_{II}}{}' P_1(\varepsilon_I)P_1(\varepsilon_{II}) = \sum_{\varepsilon_I} P_1(\varepsilon_I)P_1(E - \varepsilon_I)$$

Prime shows sum only includes terms in the expansion where $\varepsilon_{\mathrm{I}} + \varepsilon_{\mathrm{II}} = E$

Multivariable distributions with independent stochastic variables

How does the behavior of the macrostate change with increasing numbers of variables?

Role of multiple dice



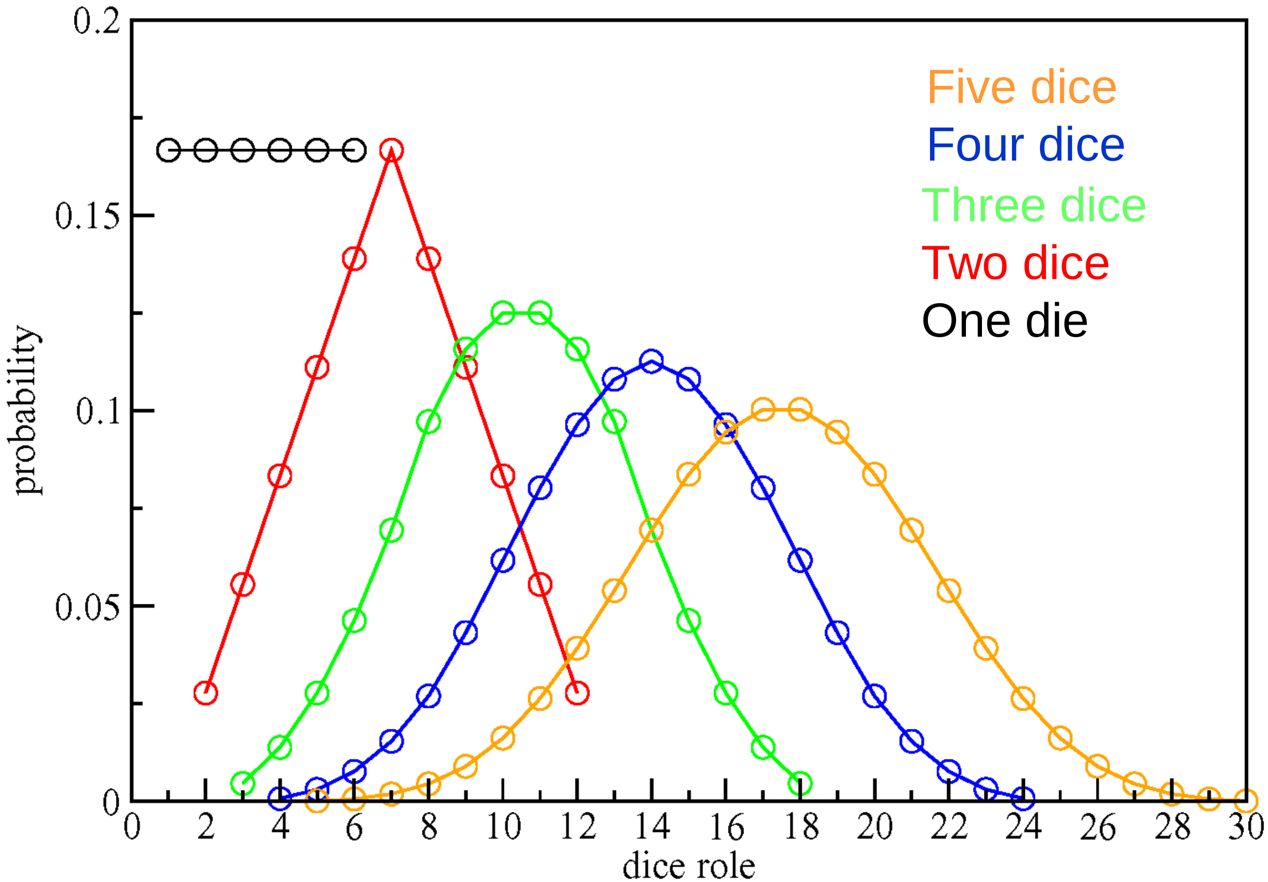
$P_N(X)$ with
$X = x_\mathrm{I} + x_\mathrm{II} + \ldots + x_N$

Note that the $P_N(X)$ distributions start to look Gaussian, even though the $P_1(x)$ are constant functions

Central Limit Theorem: A general property of probability distributions for large numbers of stochastic variables

In the limit of large N, any "reasonable" one-variable distribution function $P_1(x)$ gives a Gaussian distribution for $P_N(X)$ where $X = x_1 + x_2 + \ldots + x_N$

$$P_N(X) \xrightarrow[N\to\infty]{} \frac{1}{\sqrt{2\pi}\sigma_N} e^{-\left(X-\langle X\rangle\right)^2 / 2\sigma_N^2}$$



Averages for distributions with large numbers of stochastic variables

1) The average of the sum X is the sum of the averages of the individual variables:

$$\langle X \rangle = \int_{X_{\min}}^{X_{\max}} X P_N(X) dX$$

$$= \int_{X_{\min}}^{X_{\max}} \left(x_I + x_{II} + \cdots + x_N \right) P_1(x_I) P_1(x_{II}) \cdots P_1(x_N) dx_I dx_{II} \cdots dx_N$$

$$= \langle x_I \rangle + \langle x_{II} \rangle + \cdots + \langle x_N \rangle = \sum_{i=1}^{N} \langle x_i \rangle$$

Exercise: prove

2) The variance of X is the sum of the variances of the individual variables:

$$\left\langle \left(X - \langle X \rangle \right)^2 \right\rangle = \int_{X_{\min}}^{X_{\max}} \left(X - \langle X \rangle \right)^2 P_N(X) dX$$

$$= \int_{X_{\min}}^{X_{\max}} \left(x_I + x_{II} + \cdots + x_N - \langle x_I \rangle - \langle x_{II} \rangle - \cdots + \langle x_N \rangle \right)^2 P_1(x_I) P_1(x_{II}) \cdots P_1(x_N) dx_I dx_{II} \cdots dx_N$$

$$\sigma_N^2 = \sum_{j=1}^{N} \sigma_1^2(x_j)$$

Exercise: prove

Properties of multi-variable probability distributions

If the distribution functions for individual variables are identical:

- The average of the *N*-variable sum *X* is the sum of the 1-variable averages

$$\langle X \rangle = N \langle x_1 \rangle$$

- The variance of the *N*-variable sum is the sum of the 1-variable variances

$$\sigma_N^2 = N\sigma_1^2 \qquad \sigma_N = \sqrt{N}\sigma_1 \Rightarrow \sigma_N \propto \sqrt{N}$$

The spread of an *N* variable distribution relative to the mean becomes smaller as the number of variables increases!

$$\therefore \frac{\sigma_N}{\langle X \rangle} = \frac{\sqrt{N}\sigma_1}{N\langle x_1 \rangle} \propto \frac{1}{\sqrt{N}}$$

Convolution of continuous variable distribution functions

How do we determine the probability of getting a particular value of X from the combining continuous variables x_1 and x_2?

A similar expression for continuous variables:

$$P_2(X)dX = \iint_{X < x_I + x_{II} < X + dX} P_2(x_I, x_{II}) dx_I dx_{II} = \iint_{X < x_I + x_{II} < X + dX} P_1(x_I) P_1(x_{II}) dx_I dx_{II}$$

Subscript shows limits on the range of the integrals so that always $x_{\mathrm{I}} + x_{\mathrm{II}} = X$

The constraint on the limits of the integral makes its solution difficult.

The integral can be simplified with the use of the Dirac delta-function

$$\delta(a) = \begin{cases} 0 & \text{if} \quad a \neq 0 \\ 1 & \text{if} \quad a = 0 \end{cases}$$

In the present case, the Dirac-delta function with the argument $a = (x_1 + x_2 - X)$ is :

$$\delta(x_I + x_{II} - X) = \begin{cases} 0 & \text{if} \quad x_I + x_{II} - X \neq 0 \\ 1 & \text{if} \quad x_I + x_{II} - X = 0 \end{cases}$$

Delta function allows us to calculate the integral over x_{II}:

$$P_2(X)dX = \int_{x_{I,\min}}^{x_{I,\max}} \int_{x_{II,\min}}^{x_{II,\max}} \delta(x_I + x_{II} - X) P_1(x_I) P_1(x_{II}) dx_I dx_{II}$$

Integral over x_{II} is non-zero only when $x_{II} = X - x_I$

Final result for the convolution of independent variables:

$$P_2(X)dX = \int_{x_{I,\min}}^{x_{I,\max}} P_1(x_I) P_1(X - x_I) dx_I$$

By substituting the form of P_1 into the integral, it can be solved to give P_2.

See Chapter 5 for examples of applying this relation to Gaussian distributions. See also next lectures.

By convoluting P_2 with P_1, we can obtain P_3. Repeating the process gives P_N.