# Random processes and probability distributions

*Phys 750 Lecture 16*

# Random processes

- Many physical processes are random in character: e.g.,
  - nuclear decay (Poisson distributed event count)

$$P(k,\tau) = \frac{e^{-\lambda\tau}(\lambda\tau)}{k!}$$

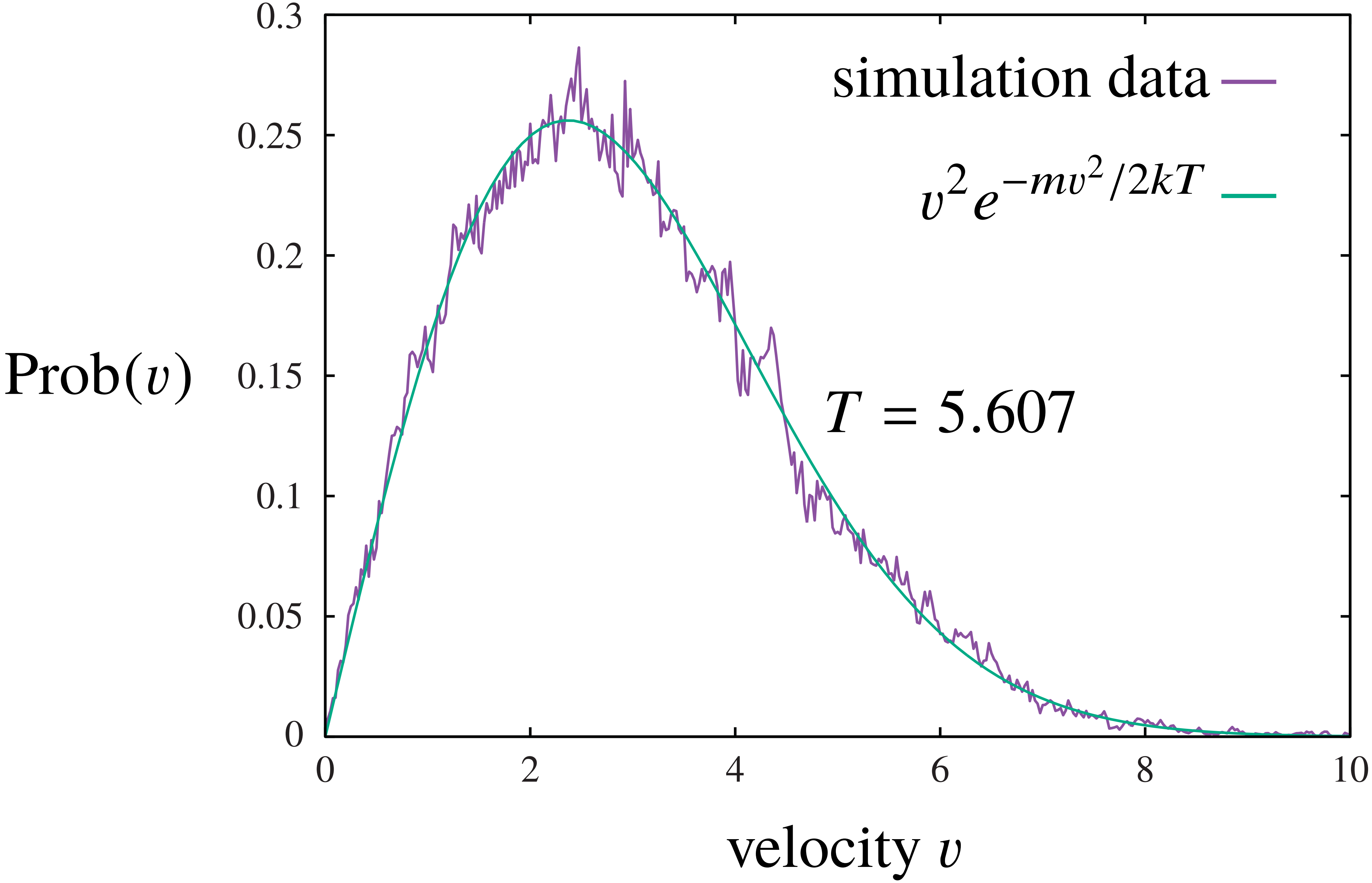
  - motion of “thermalized” interacting particles (Maxwell-Boltzmann speed profile)

$$f(v) \sim v^2 \exp\left(-\frac{mv^2}{2kT}\right)$$

# Random processes

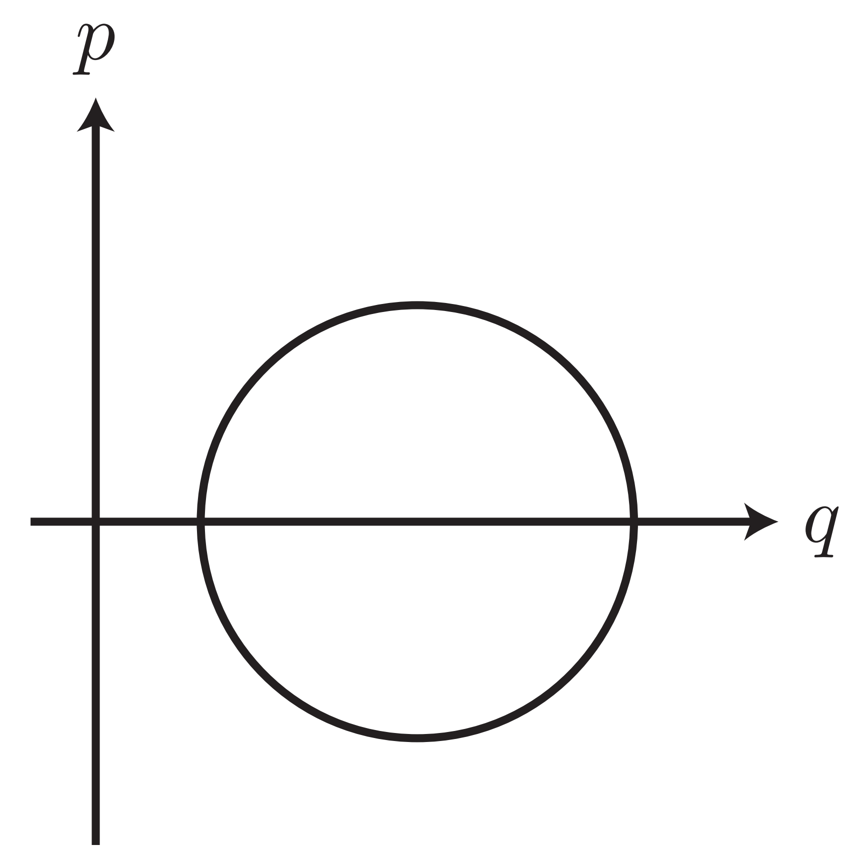
- Probabilistic descriptions often arise from the complex behaviour of many interacting degrees of freedom
- E.g., Temperature is an emergent phenomenon:
  - it’s a collective property of a large number of interacting particles
  - particles exchange energy and establish a MB distribution, characterized by a single parameter $T$

# Lennard-Jones gas



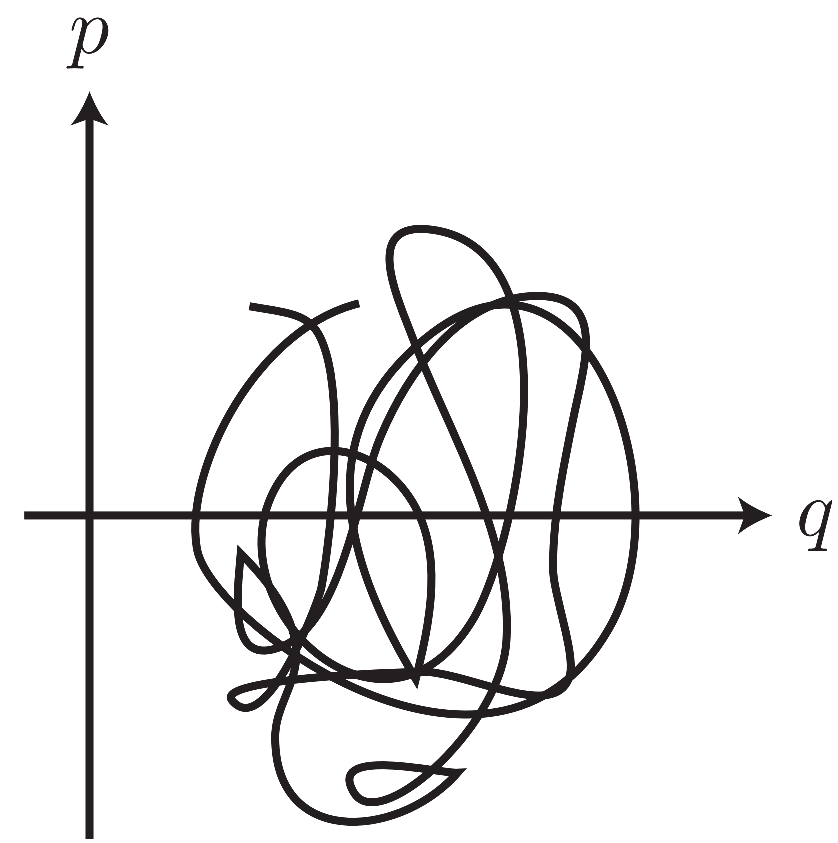
# Ergodicity

- For a large collection of non-interacting oscillators, the phase space trajectory of each individual oscillator is a closed circle



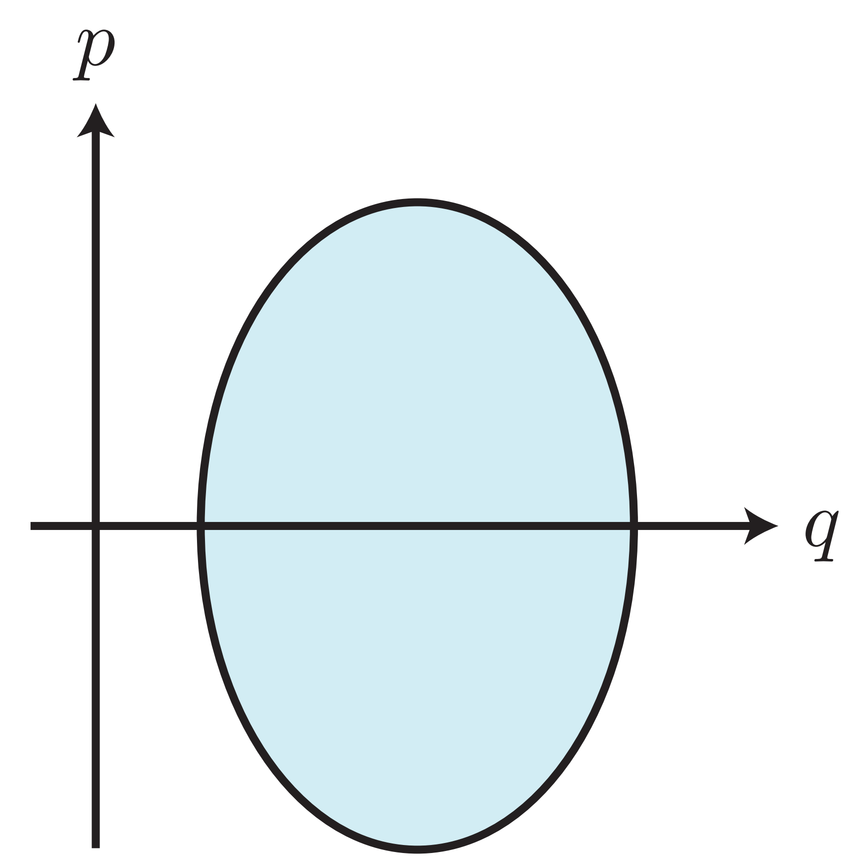
# Ergodicity

- For a large collection of non-interacting oscillators, the phase space trajectory of each individual oscillator is a closed circle
- As we turn on interactions, the trajectories are perturbed



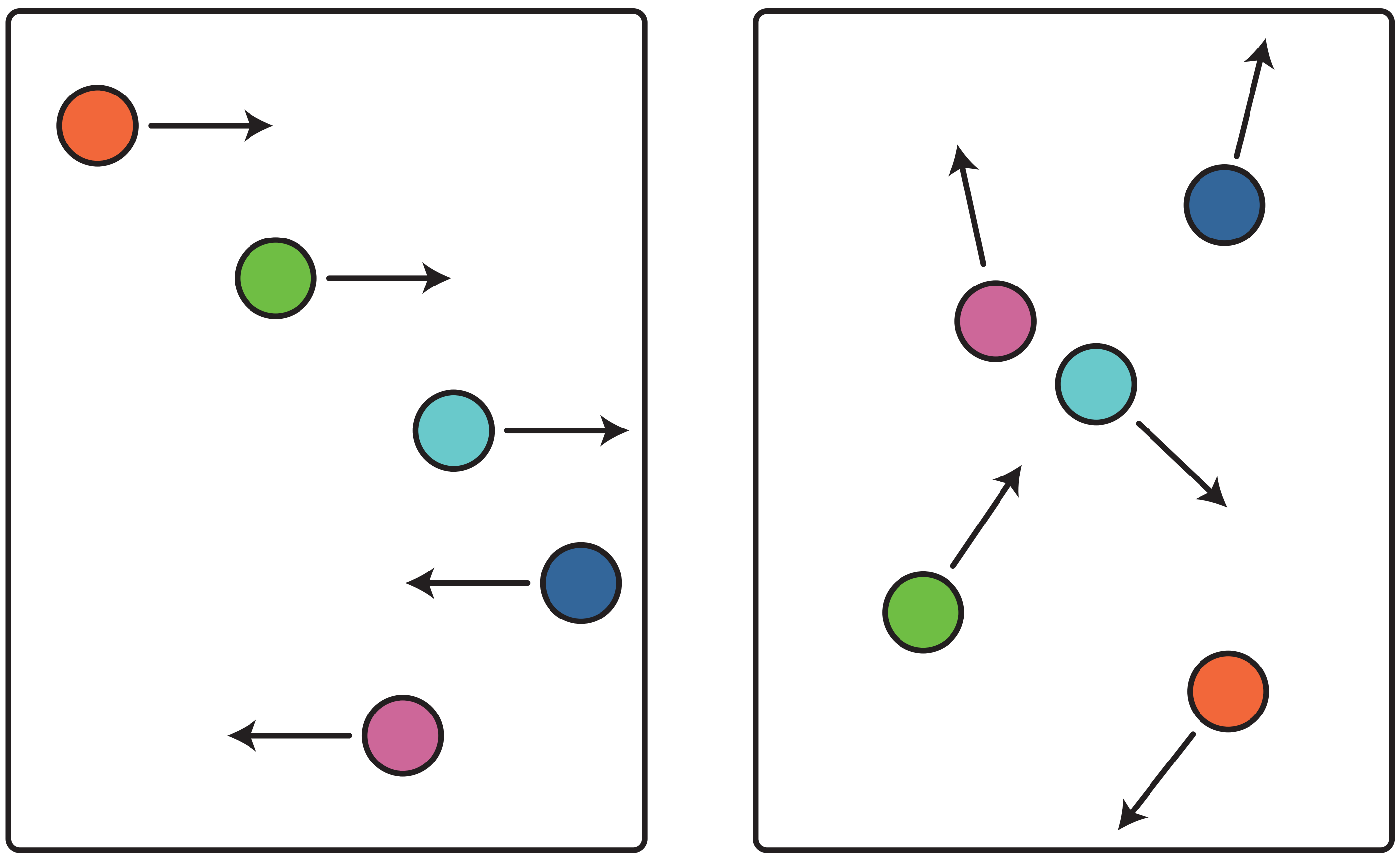
# Ergodicity

- For a large collection of non-interacting oscillators, the phase space trajectory of each individual oscillator is a closed circle
- As we turn on interactions, the trajectories are perturbed
- The trajectory density averaged over all oscillators can be understood as a probability density



# Ergodicity

- Consider an idealized billiards system. . .



non-ergodic ergodic

# Statistical physics

- Ergodicity is the basis for statistical physics
- We assume that all states in the phase space are accessible and have equal weight
- In a system at fixed temperature, states are visited according to the Boltzmann factor:

$$Z = \sum_n e^{-E_n/kT}$$

- Stochastic process with transition probabilities given by

$$P_{n\to m} \sim e^{-(E_n - E_m)/kT}$$

# Some subtleties

- Can we prove ergodicity? Rarely
- What if the dynamics are ergodic but the timescale for traversing the phase space is slow? Glassiness
- What if the dynamics are ergodic within distinct regions of the phase space that are only tenuously connected? Rare tunnelling events
- How do we model such processes on a digital state machine that is <u>purely deterministic</u>? Pseudo-randomness

# Random numbers

- Misleading terminology
- Is 23 a random number?
- In what sense could it be random?

# Random numbers

- What is a random number?
  - Really no such thing
  - Loose term of art referring to a sequence of independent numbers drawn randomly from some distribution
  - Typically these are integer or real values uniformly distributed in some finite range

# Random numbers

- An infinite sequence of digits:

$$991819562115852634258707693113278271779534707784192\cdots$$

- Is it random? (Humans are terrible at judging)
- Each of the digits $0-9$ occurs $\frac{1}{10}$ of the time
- Each pair of two successive digits occurs $\frac{1}{100}$ of the time

# Random numbers

- Consider the first one million digits of the sequence:
  - digit counts are distributed around the average
  - every pattern is equally probable

$$P(000000\cdots) = P(193273\cdots)$$

  - digits are completely uncorrelated

given $000000\cdots0x$ , $P(x=0) = \frac{1}{10}$

# Random numbers

- How do we generate sequences of random numbers?
- Strictly speaking, this isn't possible on a deterministic computer using finite arithmetic
- Nonetheless, it may be possible to construct long sequences with the appearance of randomness
- Probably okay if the relationship between numbers has no physical significance

# Linear congruential generator

- Want a random sequence of real numbers $(U_n) \in [0, 1)$
- Popular strategy:
  - use fractions $U_n = X_n/m$ built from the sequence of integers $(X_n) \in \{0, 1, 2, \ldots, m-1\}$
  - linear congruence scheme (Lehmer 1948)

$$X_{n+1} = (aX_n + c) \mod m$$

multiplier ($a$), increment ($c$), modulus ($m$)

# Linear congruential generator

- Recursion builds off an initial "seed" value, $X_0$

$$X_0$$

$$X_1 = (aX_0 + c) \bmod m$$

$$X_2 = (a[(aX_0 + c) \bmod m] + c) \bmod m$$

$$\vdots$$

- Requires careful choice of parameters:

$$(X_n) = 7, 6, 9, 0, 7, 6, 9, 0, \ldots \qquad X_0 = a = c = 7, \quad m = 10$$

very non-random, period 4

# Linear congruential generator

- Any generator of the form $X_{n+1} = F(X_n)$ taking $m$ distinct values must be periodic with period $P \leq m$
- For a linear congruential generator, one can show that the period is maximum if

  - $c$ is relatively prime to $m$;
  - $b = a - 1$ is a multiple of $p$, for very prime $p$ dividing $m$
  - $b$ is a multiple of 4, if $m$ is a multiple of 4.

# Linear congruential generator

- For the sake of efficiency, specialize to the case $m = 2^{32}$
- Advantages:
  - each integer fits into exactly one computer word
  - the modulus (an expensive division operation) is automatically handled in hardware by overflow
- A specialized version of the maximum period rule:

$$c = 1$$
$$a \equiv 5 \bmod 8$$

# Linear congruential generator

- A long repeating cycle does not imply randomness:

$$(X_n) = 0, 1, 2, \ldots, m-1 \quad (a = c = 1)$$

- We require weak correlation between elements,

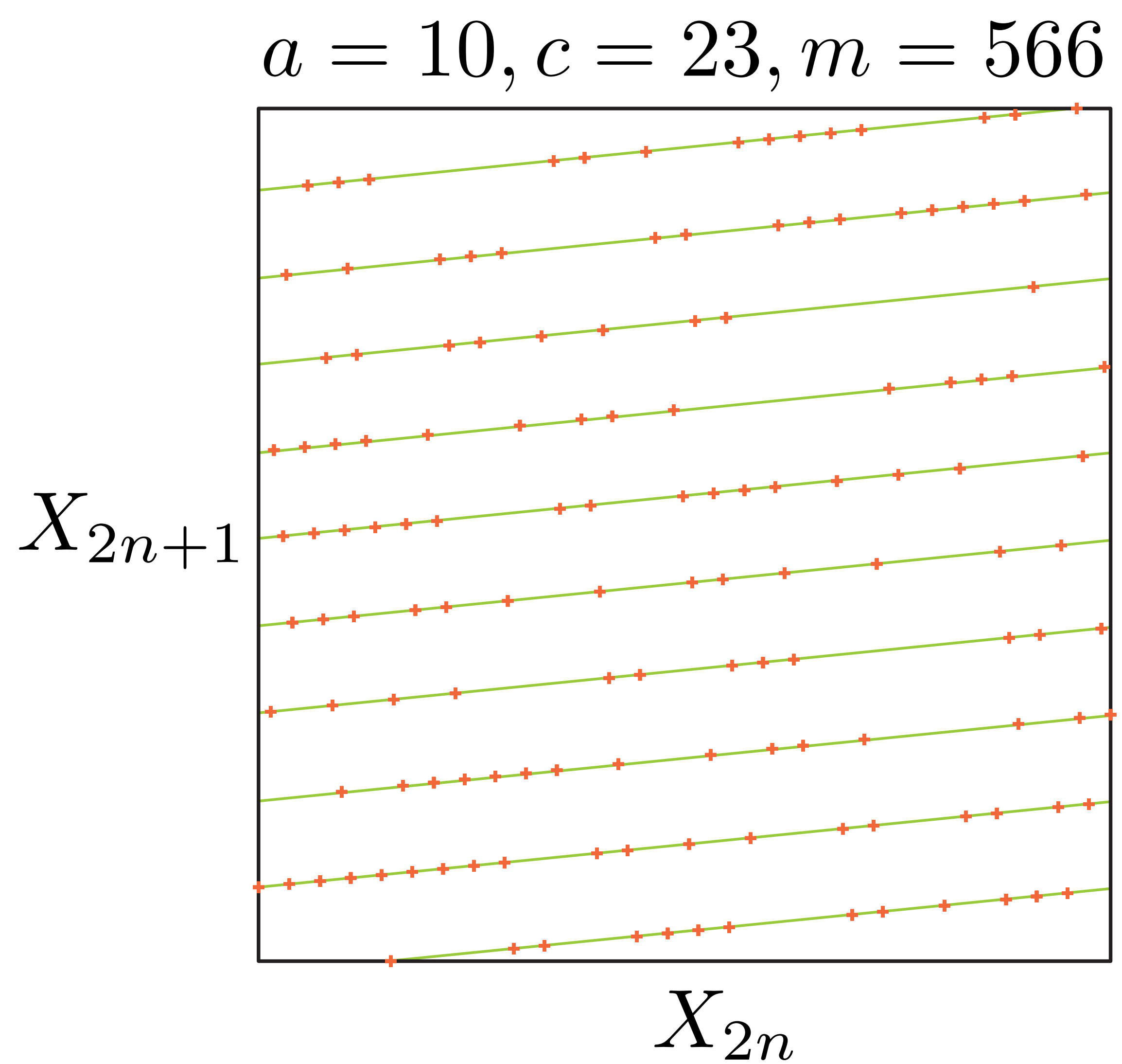
$$\langle X_j X_k \rangle \approx \langle X_j \rangle \langle X_k \rangle \ (j \neq k),$$

especially when $|j - k|$ is small

- Judged empirically via statistical tests (Die Hard)

# Linear congruential generator

- Some other considerations:
  - best if $a/m$ is not too small
  - least significant bits are more highly correlated
  - complete orbits lie in hyperplanes (Marsaglia 1968):




$$(X_0, X_1, \ldots, X_{q-1}), (X_q, X_{q+1}, \ldots, X_{2q-1}), \ldots$$

# Probability distributions

- Suppose there are $N$ discrete events occur with probabilities $p_1, p_2, \ldots, p_N$
- Since something must happen, the total sum is

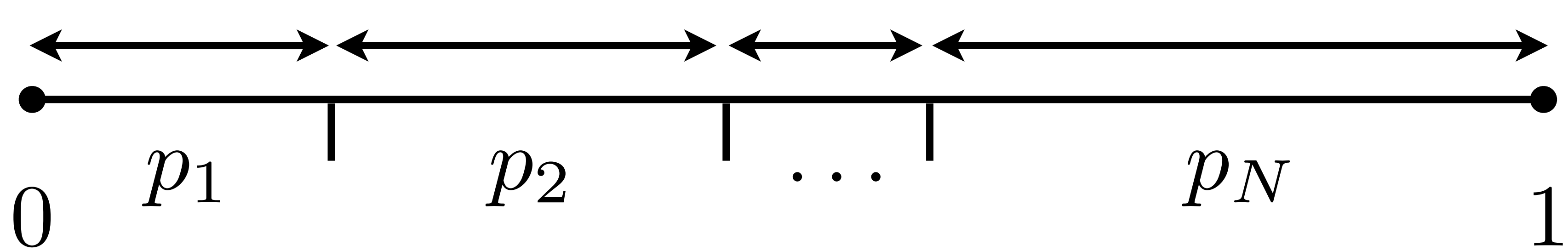
$$p_1 + p_2 + \cdots + p_N = 1$$

- Given a randomly generated number $\xi \in [0, 1]$, how can we select one of the $N$ events?

# Probability distributions

- Each even $i$ occupies a width $p_i$ in the interval:

$$
\begin{aligned}
1 &\leftarrow\ 0 < \xi < p_1 \\
2 &\leftarrow\ p_1 < \xi < p_1 + p_2 \\
3 &\leftarrow\ p_1 + p_2 < \xi < p_1 + p_2 + p_3 \\
&\ \vdots \\
N &\leftarrow\ p_1 + \cdots + p_{N-1} < \xi < 1
\end{aligned}
$$



# Probability distributions

- The relevant quantity is the cumulative probability,

$$P_i = \sum_{j=1}^{i} p_j$$

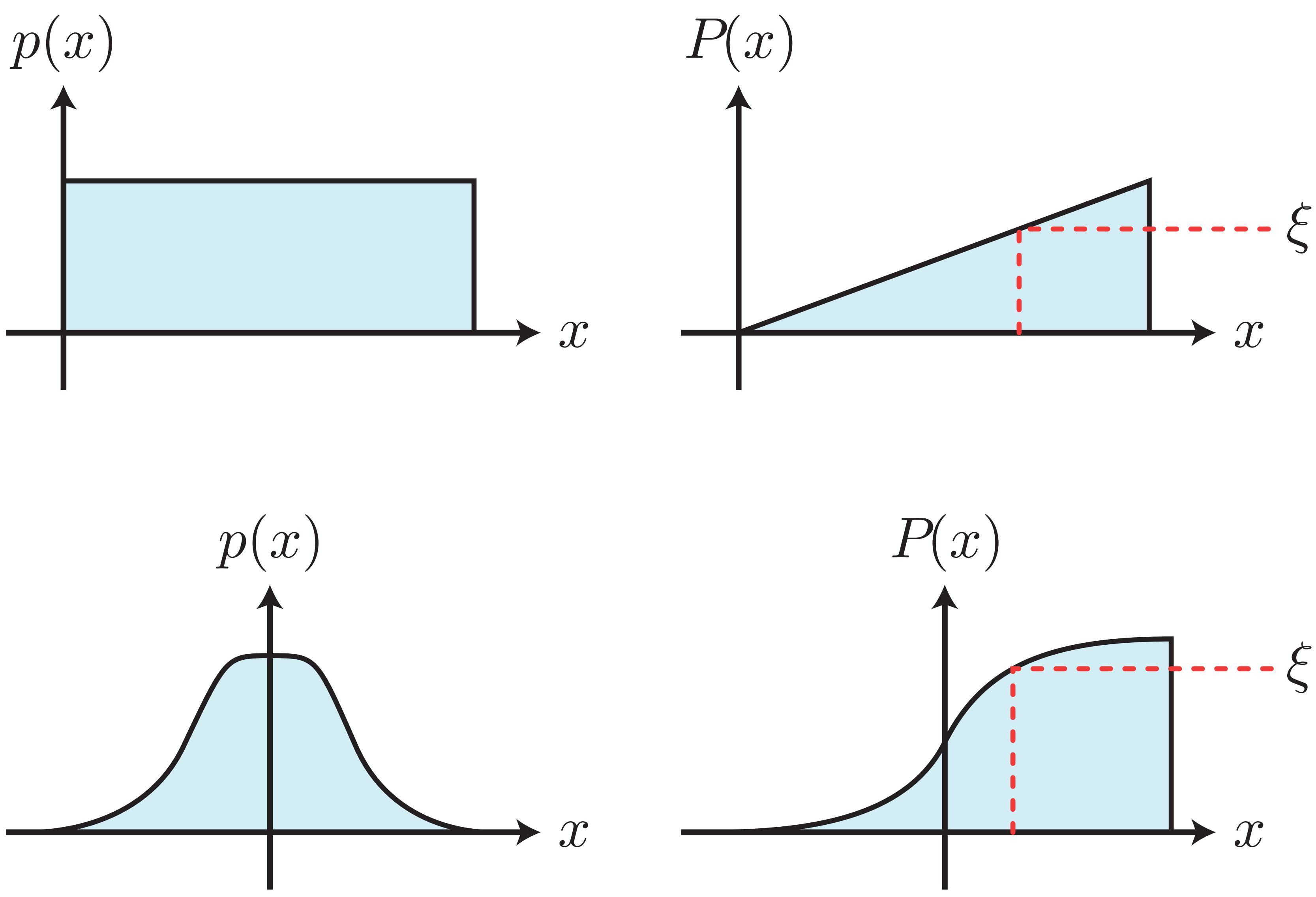
- Similarly, for continuous distributions, we construct a cumulative probability distribution

$$P(x) = \int_{-\infty}^{x} dy\, p(y)$$

from the probability density $p(x)$

# Probability distributions

- Sampling via $x \leftarrow P^{-1}([0,1])$



# Inverse transform method

- Example: $p(x) = \begin{cases} (1/\lambda)e^{-x/\lambda} & \text{if } 0 \le x < \infty \\ 0 & \text{if } x < 0 \end{cases}$

- Back map gives $\xi \to P(x) = \int_0^x dy\, p(y) = 1 - e^{-x/\lambda}$

- By inversion, we see that $x \leftarrow -\lambda \ln(1-\xi)$ with $\xi$ drawn uniformly from $[0,1]$ is equivalent to $x$ drawn from the nonuniform distribution $p(x)$

# Inverse transform method

- What about cases where no analytic inverse exists?
- In some cases, related multivariate distributions are invertible: e.g. Gaussian distribution

$$p(x) = \frac{e^{-x^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}$$

- Consider the product

$$p(x, y) = p(x)p(y) = \frac{e^{-(x^2+y^2)/2\sigma^2}}{2\pi\sigma^2}$$

# Inverse transform method

- From radial coordinates $r = \sqrt{x^2 + y^2}$

$$\theta = \tan^{-1}(y/x)$$

we can further transform $\rho = r^2/2$ so that

$$x = \sqrt{2\rho}\cos\theta$$

$$p(\rho, \theta) d\rho\, d\theta = \frac{1}{2\pi} e^{-\rho} d\rho\, d\theta$$

$$y = \sqrt{2\rho}\sin\theta$$

- Sampling is now possible with two random variables:

$$x \leftarrow \sqrt{-2\ln(1-\xi_1)}\cos(2\pi\xi_2)$$

**(Box-Muller)**

# Rejection method

- When $P(x)$ is not easily invertible and no other tricks can be applied, try the following <u>rejection method</u>:
  - Generate a sequence $(x_1, x_2, x_3, \cdots)$ with the elements drawn uniformly from $[x_{\rm min}, x_{\rm max}]$
  - Generate a sequence $(\xi_1, \xi_2, \xi_3, \cdots)$ with the elements drawn uniformly from $[0, p_{\rm max}]$
  - Discard elements from the first sequence if $p(x_i) < \xi_i$

# Rejection method

- Works by throwing away results that are rare
- Some limitations: the probability distribution must be bounded and have a finite range