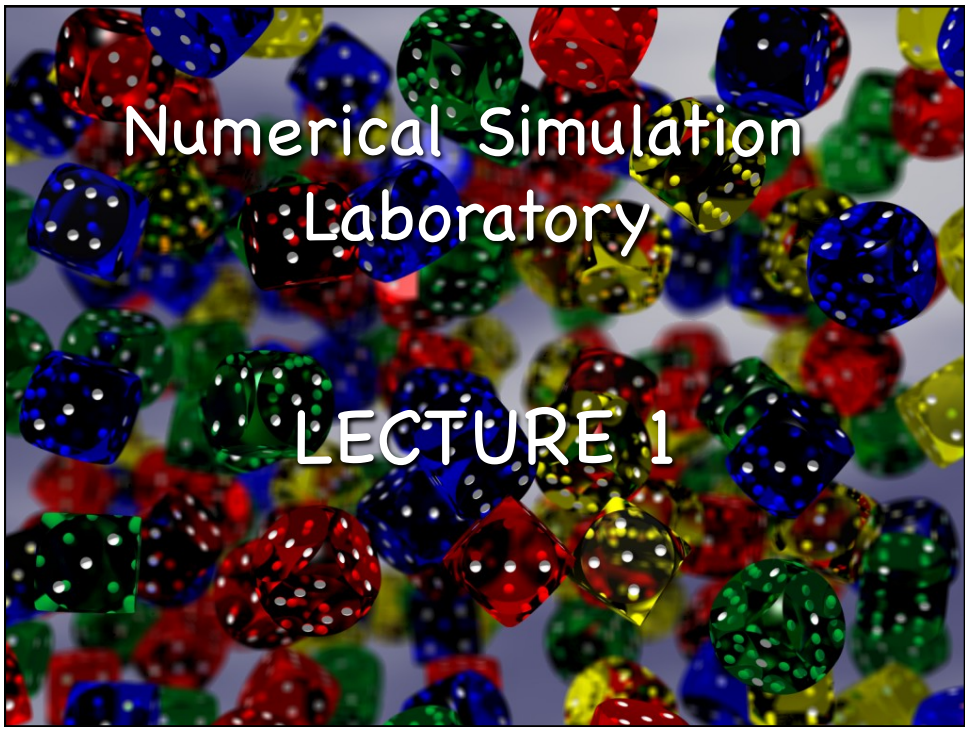


Lecture 1 - outline

- Some basic of probability theory
- The Central Limit theorems
- Random numbers and pseudo-random number generators
- Sampling of random variables
- Transformation of random variables

Numerical simulation & stochastic approaches

- The extraordinary increase of computer power in the last decades encourages physicists, chemists, biologists, economists, and engineers to model and simulate numerically tremendously complex phenomena, in order to answer basic scientific questions, industrial needs, and societal requirements for control, profit and risk evaluation.
- Stochastic approaches appear to be useful, and sometimes mandatory, in the following 2 contexts:

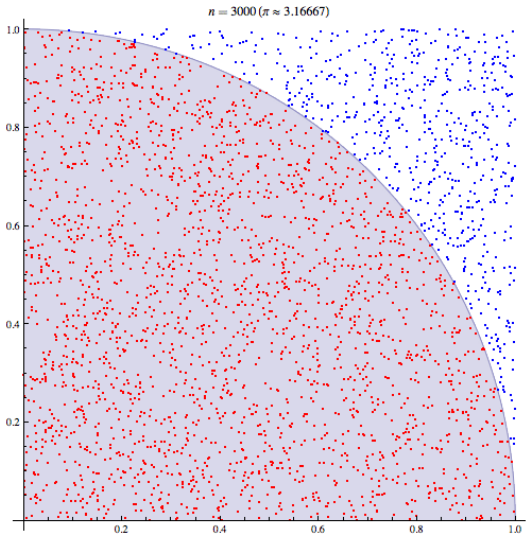
1. First, one cannot expect that very complex phenomena lead to perfectly calibrated mathematical models, or to perfect mathematical models, or even to theories intrinsically probabilistic (e.g. Statistical & Quantum Mechanics), so that uncertainties or stochastic components are involved in the equations.
2. Second, stochastic numerical methods allow one to solve deterministic problems, of which characteristics render classical deterministic methods of resolution intractable or inaccurate, provided that the solutions can be represented in terms of probability distributions of random variables or stochastic processes.

What is Monte Carlo?

- The name Monte Carlo was applied to a class of mathematical methods first by scientists working on the development of nuclear weapons in Los Alamos in the 1940s (see supplementary material – introductory lecture).
- The essence of the method is the invention of statistical/random sampling methods whose outcome can be used to study interesting phenomena.
- While there is no essential link to computers, the effectiveness of random sampling as a serious scientific pursuit is enormously enhanced by the availability of modern computers.
- It is interesting, and may strike some as remarkable, that carrying out random sampling will produce anything worthwhile… let me start with a simple example …

- Consider a circle and its circumscribed square. The ratio of the area of the circle to the area of the square is π/4. It is plausible that if points were placed at random in the square, a fraction π/4 would also lie inside the circle.
- It is also possible to program a computer to generate random pairs of Cartesian coordinates to represent random points in the square and count the fraction that lie in the circle. This fraction as determined from many experiments should be close to π/4, and the fraction would be called an estimate for π/4.
- The example illustrates that random sampling may be used to solve a mathematical problem, in this case, evaluation of a definite integral,

$$I = \int_0^1 \int_0^{\sqrt{1-x^2}} dx\,dy$$

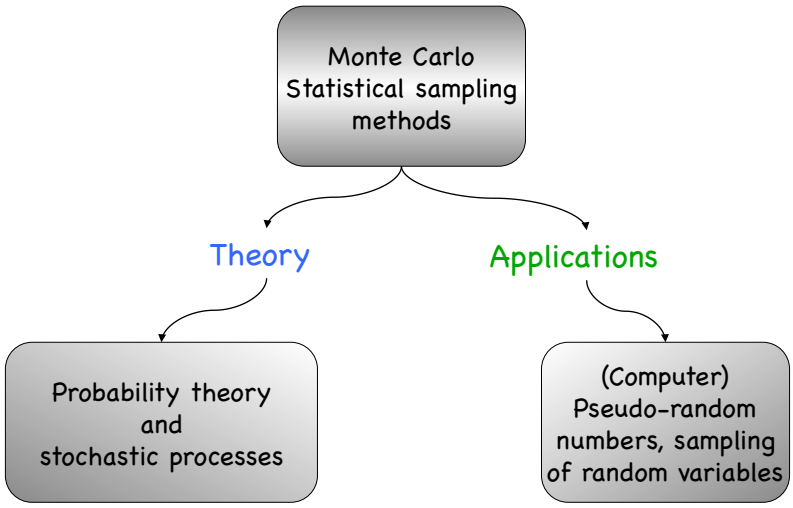


- The answers obtained are statistical in nature and subject to the laws of chance. This aspect of Monte Carlo is a drawback, but not a fatal one since one can determine how accurate the answer is, and obtain a more accurate answer, if needed, by conducting more experiments.
- Sometimes, in spite of the random character of the answer, it is the most accurate answer that can be obtained for a given investment of computer time.
- Typically, Monte Carlo methods are in fact computationally effective, compared with deterministic methods, when treating problems in high dimensional spaces

Monte Carlo methods

- A definition of a Monte Carlo method would be one that involves deliberate use of random numbers in a calculation that has the structure of a stochastic process.
- By stochastic process we mean a sequence of states whose evolution is determined by random events.
- In a computer, these are generated by (pseudo) random numbers.
- A distinction is sometimes made between simulation and Monte Carlo: In this view, simulation is a rather direct transcription into computing terms of a natural stochastic process or of a complex system. Monte Carlo, by contrast, is the solution by probabilistic methods of non-probabilistic problems (as in the example of π).
- The distinction is somewhat useful, but often impossible to maintain. We will deliberately use Monte Carlo and simulation on both kind of stochastic approaches.

What is at the basis of Monte Carlo methods?



Events, probability, sample space

9

- Probability is a quantification of our expectation of the outcome of an experiment
- Event = the outcome of an experiment involving randomness
- Suppose that one possible outcome of an experiment is A. Then, the probability of A occurring is P(A) if, out of N identical experiments, we expect that N·P(A) will result in the outcome A. As N becomes very large ($N \to \infty$), the fraction of experiments which result in A will approach P(A).
- The concept of a sample space is often useful for obtaining relations between probabilities and for analyzing experiments. The collections of all possible outcomes of an experiment is called sample space: a sample space of an experiment is a set, Ω, of elements such that any outcome of the experiment corresponds to one or more elements of the set
- An event is a subset of a sample space Ω of an experiment.

Simple and composite events

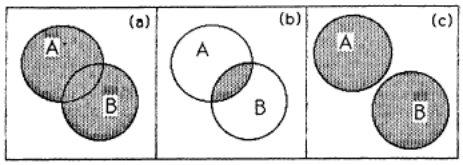
10

- There are certain events among these that are indecomposable and are called simple events
- Example: rolling a die (finite and discrete sample space)
  - The sample space is Ω={1,2,3,4,5,6}, the collection of all possible outcomes of rolling a die
  - The simple events are {1}, {2}, {3}, {4}, {5}, {6}
  - Other possible (composite) events are:
    A_1: {the result is an even number} A_2:{the result is larger than 3}

Set-theory concepts…

11

- Some ideas from set theory are useful:
  (Fig.a) The union of two events A_1 and A_2 is denoted $A_1 \cup A_2$. $A_1 \cup A_2$ is the set of all simple events belonging to A_1 or A_2 or both.
  (Fig.b) The intersection of two events is denoted $A_1 \cap A_2$. $A_1 \cap A_2$ is the set of all simple events belonging to both A_1 and A_2.
  (Fig.c) If the events A_1 and A_2 are mutually exclusive, then $A_1 \cap A_2 = \varnothing$ where $\varnothing$ is the empty set



- The previously introduced $A_1=\{2,4,6\}$ and $A_2=\{4,5,6\}$ are unions of the simple events: $A_1=\{2\}\cup\{4\}\cup\{6\}$ and $A_2=\{4\}\cup\{5\}\cup\{6\}$
  for both of them we can also define their complement, i.e. A'_1:"rolling an odd number" and A'_2:"rolling a number smaller or equal to 3".
  (In a different notation $A'_1=\Omega\backslash A_1$ and $A'_2=\Omega\backslash A_2$)

… and the joint probability

12

- The event "rolling an even number larger than 3" is the simultaneous occurrence of A_1 and A_2 , and is given by the intersection of the sets $A_1 \cap A_2=\{4,6\}$
- The event "rolling a number larger than 3 or rolling an even number", which is the union of A_1 and A_2 , $A_1 \cup A_2=\{2,4,5,6\}$
- We shall let $P(A)$ denote the probability that event A is the outcome of an experiment ($P(\varnothing)=0$, $P(\Omega)=1$)
- We shall let $P(A_1 \cap A_2)$ denote the probability of the simultaneous occurrence of A_1 and A_2 , we can call $P(A_1 \cap A_2)$ a joint probability
- Finally we shall let $P(A_1 \cup A_2)$ denote the probability that event A_1 or event A_2 or both occur as the outcome of an experiment…

Independent events

13

- … then the probability $P(A_1 \cup A_2)$ may be written

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

In fact, in writing $P(A_1) + P(A_2)$ we account twice the simple events contained in $A_1 \cap A_2$; therefore we have to subtract $P(A_1 \cap A_2)$

- If the two events A_1 and A_2 are mutually exclusive, then they have no simple events in common and

$$P(A_1 \cup A_2) = P(A_1) + P(A_2)$$

- If events A_1, A_2, A_3, …, A_m are mutually exclusive and exhaustive, then $A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_m = \Omega$ and the m events form a partition of the sample space Ω into m subsets, moreover $P(A_1) + P(A_2) + P(A_3) + \ldots + P(A_m) = 1$
- The events A_1 and A_2 are independent if and only if

$$P(A_1 \cap A_2) = P(A_1) \times P(A_2)$$

Independent events are not mutually exclusive events. They are completely different concepts.

Conditional probability

14

- The conditional probability $P(A_2|A_1)$ gives us the probability that event A_2 occurs as the result of an experiment if A_1 also occurs. $P(A_2|A_1)$ is defined by the equation

conditional → $P(A_2|A_1) = \dfrac{P(A_1 \cap A_2)}{P(A_1)}$ ← joint, ← marginal

- Since $P(A_1 \cap A_2) = P(A_2 \cap A_1)$ we find also that

$$P(A_2|A_1)P(A_1) = P(A_1|A_2)P(A_2)$$

- From $P(A_1 \cap A_2) = P(A_1)P(A_2)$, valid if A_1 and A_2 are independent, we see that if A_1 and A_2 are independent, then

$$P(A_2|A_1) = P(A_2)$$

15

Role of probability theory

- IMPORTANT: in order to find out the probabilities for all events, we first have to know what probabilities to assign to the simple events, which cannot be deduced within the theory of probability
- Probabilities of the simple events are input into a probabilistic treatment of a given problem, we cannot deduce them within the theory of probability: they are either obtained from empirical knowledge, by intuition or on the basis of some different theoretical description of the problem
- Probability theory is only concerned with rules to manipulate these basic probabilities in order to calculate the probabilities of the more complex events under consideration
- In the case of a die our intuition and our hope for a fair game lead us to assign a probability of p=1/6 for all six outcomes of rolling a single die once (but think about a fake die…!). From this we can deduce: $p(A_1)= p(A_2)=1/2$, $p(A_1 \cap A_2)=1/3$, $p(A_1 \cup A_2)=2/3$ and, of course, $p(\Omega)=1$

16

Probability spaces

- What is the mathematical structure to deal with events and their probabilities when situations are not describable within the simple framework of a discrete and finite sample space? The axiomatic answer to the previous question was formulated by Kolmogorov in 1931
- Definition: The sample space is a probability space, i.e. a measure space $(\Omega,\mathcal{F},\mu)$, given through a set Ω, a σ-algebra $\mathcal{F}$ of subset of Ω and a measure μ on $\mathcal{F}$. An event is a set $A\in\mathcal{F}$ and the probability for this event is the measure $\mu(A)$. The measure has the property: $\mu(\Omega)=1$
- A σ-algebra has the following properties that correspond to the intuitively required properties of events, as we discussed them for the rolling die:
  - $\Omega\in\mathcal{F}$ and $\varnothing\in\mathcal{F}$ where $\varnothing$ is the empty set
  - $A\in\mathcal{F} \Rightarrow \Omega\backslash A=A' \in\mathcal{F}$
  - $A,B\in\mathcal{F} \Rightarrow A\cap B\in\mathcal{F}$
  - $A_i\in\mathcal{F}$, i=1,2,… $\Rightarrow \cup_{i=1,\dots}A_i\in\mathcal{F}$;
    if furthermore $A_i\cap A_j= \varnothing$ for $i\neq j$ then $\mu(\cup_{i=1,\dots}A_i)=\Sigma_{i=1,\dots}\mu(A_i)$

17

- The collection of simple sets, together with the empty set and the whole space, generates the σ-algebra through the operations of building the complement, finite intersections and countable unions
- Again this does not specify the simple events uniquely. Their choice, and the corresponding σ-algebra, is an input into the theory and depends on our prior information about the problem that we want to treat probabilistically
- The simplest σ-algebra is $\mathcal{F}_0=\{\varnothing,\Omega\}$, where all we can tell is that the event Ω occurred
- If we want to distinguish one subset A∈Ω in our "experiment", we can define the σ-algebra $\mathcal{F}_1=\{\varnothing,A,A',\Omega\}$. To complete the construction of a probability space, however, we also have to be able to assign some probability to the set A.
- Again, note that in the last example $\mu(A)$ [and $\mu(A')=1-\mu(A)$] can be any number in [0,1], it depends on what you are modelling

18

Random Variables

- Definition: A random (stochastic) variable, f, is a measurable function on the probability space $(\Omega,\mathcal{F},\mu)$.
- The expectation value of the random variable f is defined as

$$\langle f \rangle \equiv E[f] := \int_\Omega f \, d\mu$$

This means that f is in $L^1(\Omega,\mathcal{F},\mu)$, the space of μ-integrable functions on Ω.

- If f is also in $L^n(\Omega,\mathcal{F},\mu)$, we can define the n-th moment of f as

$$\langle f^n \rangle \equiv E[f^n] := \int_\Omega f^n \, d\mu$$

- Definition: A measure μ is called absolutely continuous with respect to another measure ν (written $\mu \ll \nu$) if for all A∈$\mathcal{F}$ we have $\nu(A)=0 \Rightarrow \mu(A)=0$. μ and ν are equivalent when $\mu \ll \nu$ and $\nu \ll \mu$ Equivalent measures have the same sets (events) of measure zero
- Theorem (Radon-Nikodym): If $\mu \ll \nu$, then there exists an $\mathcal{F}$-measurable function p, such that for every A∈$\mathcal{F}$

$$\mu(A) = \int_A p \, d\nu$$

19

Probability density

- The function $p=d\mu/d\nu$ is called the Radon-Nikodym derivative of μ with respect to ν. If μ and ν are equivalent we have $\frac{d\mu}{d\nu} = \left(\frac{d\nu}{d\mu}\right)^{-1}$
- The probability measures we will generally encounter in the applications we discuss in these lectures are all absolutely continuous with respect to the Lebesgue measure. We will be able to write them as $d\mu(x)=p(x)dx$ and often we will use the so called probability density p, or probability distribution function (pdf), to denote the relevant probability measure.
- This is the description usually adopted in physical applications. Therefore, in such cases, the probabilistic structure of $(\Omega,\mathcal{F},\mu)$ is translated into the probabilistic structure $(\mathbb{R},\mathcal{B},p(x)dx)$ where $\mathcal{B}$ is the Borel algebra, i.e. the σ-algebra generated by the open intervals of $\mathbb{R}$ (or $\mathbb{R}^n$)
- To calculate, for instance, the expectation value of a random variable f we will evaluate

$$\langle f \rangle \equiv \bar{f} \equiv E[f] := \int_\Omega f\, d\mu = \int_{\mathbb{R}} f(x)p(x)\,dx \qquad \left(E[f] := \sum_{disc.\,\nu} p_\nu f_\nu\right)$$

20

- The n^{th} moment of the random variable f is translated into

$$E[f^n] := \int_{\mathbb{R}} f^n(x)p(x)\,dx \qquad \left(E[f^n] := \sum_{disc.\,\nu} p_\nu f_\nu^n\right)$$

- The central n^{th} moment is defined as $\int_{\mathbb{R}} \left(f(x) - \langle f \rangle\right)^n p(x)dx$

  and the most relevant is the second central moment, also called the variance, the squared deviation from the expectation:

$$\sigma^2 = \mathrm{var}[f] = \int_{\mathbb{R}} \left(f(x) - \langle f \rangle\right)^2 p(x)\,dx \qquad \left(\mathrm{var}[f] := \sum_{disc.\,\nu} p_\nu (f_\nu - \bar{f})^2\right)$$

- The square root of the variance, σ, is called standard deviation, but also Root Mean Square (RMS) deviation from the mean
- Let's now consider a set of random variables, $x_1 \ldots x_d$, in the probability space ($\mathbb{R}^d$, $\mathcal{B}$, $p(x_1,\ldots,x_d)$), $p(x_1,\ldots,x_d)$ is called the joint probability density for the random variables $x_1 \ldots x_d$.
- The covariance matrix is defined by

$$\mathrm{cov}[x_i,x_j] \equiv \sigma_{ij}^2 = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle$$

  and it is of special interest because they serve to define when two random variables are uncorrelated.

21

- Definition: two random variables x_1 and x_2 are uncorrelated if and only if $\mathrm{cov}[x_1,x_2]=0$
- Definition: two random variables x_1 and x_2 are statistically independent if and only if $p(x_1,x_2)=p_1(x_1)p_2(x_2)$, in this case

$$\left\langle x_i^n x_j^m \right\rangle = \left\langle x_i^n \right\rangle \left\langle x_j^m \right\rangle \quad \forall n,m$$

  which implies $\mathrm{cov}[x_1,x_2]=0$. Whenever two random variables are not statistically independent, there is some information about possible outcome of a measurement of one of the random variables contained in the measurement of the other.
- Suppose we know the joint distribution of the two random variables $p(x_1,x_2)$, then we can define their marginal distributions:

$$p_1(x_1) = \int_{\mathbb{R}} p(x_1,x_2)dx_2 \qquad \text{and} \qquad p_2(x_2) = \int_{\mathbb{R}} p(x_1,x_2)dx_1$$

- The conditional probability density function for x_1 given x_2 is then defined as

$$p(x_1|x_2) = p(x_1,x_2)/p_2(x_2)$$

  where $p_2(x_2)\neq 0$ is assumed. We could define $p(x_2|x_1)$ in a symmetric manner. For each fixed x_2 , $p(x_1|x_2)$ is a regular probability density.

22

Large number of random events

- Many important complex phenomena can be described as the compound effect of many small random influences (for example Brownian motion)
- Observable quantities are most often the sum of a very large number of random events (pressure exerted by a gas on a piston)
- A central question is, therefore, what the distribution of a sum of random variables will ultimately be. It can be shown that the probability density describing the distribution of outcomes of a large number of events universally approaches a Gaussian form (provided the moments of the distribution for the individual events are finite)
- This is called the Central Limit Theorem. This result shows why Gaussian distributions are so widely seen in nature
- Another result of considerable importance is the law of large numbers. The law of large numbers gives quantitative justification to the use of probabilities

The Central Limit Theorem

- We consider N statistically independent and identically distributed random variables $x_1,\ldots,x_N$, i.e.,

$$p_N(x_1,x_2,\cdots,x_N) = p(x_1)p(x_2)\cdots p(x_N)$$

we require $\langle x_1\rangle = \cdots = \langle x_N\rangle = \mu$

Furthermore, let us denote $\langle x_1^2\rangle - \langle x_1\rangle^2 = \cdots = \langle x_N^2\rangle - \langle x_N\rangle^2 = \sigma^2 < \infty$

What is the pdf of $A_N = \frac{1}{N}\sum_{n=1}^{N} x_n \left(S_N = \sum_{n=1}^{N} x_n\right)$?

It follows that:

1) $\langle A_N\rangle = \mu \quad \left(\langle S_N\rangle = \mu N\right)$

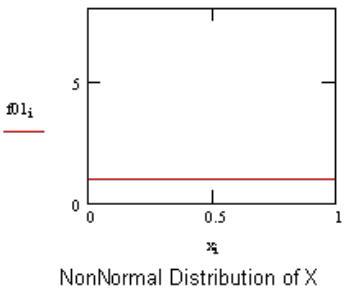
2) $\sigma^2_{A_N} = \frac{\sigma^2}{N} \quad \left(\sigma^2_{S_N} = N\sigma^2\right)$

3) $P_N(A_N) \underset{N\to\infty}{\to} \frac{1}{\sqrt{2\pi\sigma^2_{A_N}}} \exp\left[-\frac{(A_N-\mu)^2}{2\sigma^2_{A_N}}\right]$

$\left(P_N(S_N) \underset{N\to\infty}{\to} \frac{1}{\sqrt{2\pi\sigma^2_{S_N}}} \exp\left[-\frac{(S_N-\mu N)^2}{2\sigma^2_{S_N}}\right]\right)$

This result lies at the heart of the ubiquitous appearance of the Gaussian distribution in statistical phenomena. Whatever the exact underlying distribution of the individual random variables is, as long as the first two moments exist, the "average" variable is asintotically Gaussian distributed (see supplementary mat. for a proof)

- Whatever the exact form of the underlying distribution of the individual random variable is, as long as the first two moments exists, the sum variable always obeys a Gaussian distribution in the large N limit



NonNormal Distribution of X

- This theorem can be generalized to variables which are not identically distributed as long as all variables have finite variance; in this case the average tends to the average of the averages, and the variance tend to the average of the variances divided by N

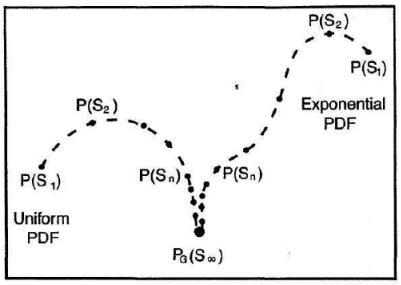
- When not-independents variables are considered, the theorem is still true if each variable is correlated with a finite number of variables in the sequence

- If N is large, the variable A_N (S_N) practically assumes the value $\langle x\rangle$, this is the law of large numbers (see supplementary material)

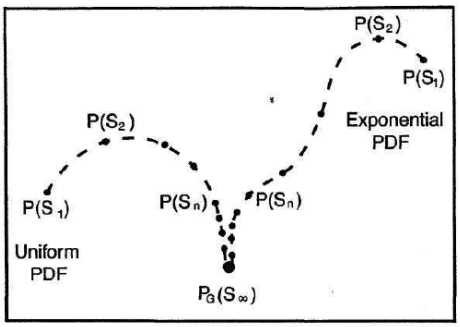
Basin of attraction

25

- The study of the Central Limit Theorem (CLT) let us to introduce the concept of the basin of attraction of a probability distribution. We focus our attention on the changes in the functional form of $P(S_N)$ that occur when N changes
- We restrict our discussion to identically distributed random variables x_i. $P(S_1)$ then coincides with $p(x_i)$. When n increases, changes its functional form and, if the hypotheses of the CLT are verified, assumes the Gaussian functional form for an asymptotically large value of N.
- The Gaussian pdf is thus an attractor (or fixed point) in the functional space of pdfs for all the pdfs that fulfill the requirements of the CLT. The set of such probability distributions constitutes the basin of attraction of the Gaussian distribution.
- In Figure we provide, as an example, a pictorial representation of the "motion" of both the uniform and exponential in the functional space of probability distributions, and sketch the convergence to the Gausssian



26



- Both S_i are obtained by summing independent identically distributed random variables. The two processes differ in their original probability distributions, indicated by their starting from different regions of the functional space.
- When N increases, both $P(S_N)$ become progressively closer to the Gaussian attractor
- The number of steps required to observe the convergence of to provides an indication of the speed of convergence of the two families of random processes
- Although the Gaussian attractor is the most important attractor in the functional space of probability distributions, other attractors also exist, and we consider them in the following.

Beyond the Central Limit Theorem

- Recall that the central limit theorem states that a sum of N independent and identically distributed random variables $S_N = \sum_{n=1}^{N} x_n$ obeys a Gaussian distribution in the limit N→∞, provided the first and second moments of x_n do not diverge.
- These restrictions are so mild that many distributions belong to the domain of attraction of the Gaussian.
- However, not all! You known that an exception is the Cauchy distribution,

$$p_C(x) = \frac{\Gamma}{\pi}\frac{1}{x^2 + \Gamma^2}$$

whose second moment is infinite.

- The Cauchy distribution occurs in many physical situations, in the Ornstein-Zernike theory of critical opalescence and in the lifetime broadening of spectral lines, for instance.
- Therefore the questions arise of whether p_C could also emerge as a limiting distribution for S_N, and what the limiting distribution would look like if the random variables were distributed according to it

Levy or stable distributions

- The Cauchy distribution is just one example of a whole class of distributions which possess long, inverse-power-law tails:

$$p(x) \underset{x\to\infty}{\approx} \frac{1}{|x|^{1+\alpha}} \qquad 0 < \alpha < 2$$

- These 'broad' tails preclude the convergence to the Gaussian for N→∞, but not the existence of a limiting distribution.
- The premises for, the form and the properties of these limiting distributions were worked out in the 1930s by P. Levy, A. Khintchine and others. They are today called Levy or stable distributions.
- Definition: A probability density is called stable if it is invariant under convolution, i.e., if there are constants a > 0 and b such that

$$p(a_1 x + b_1) * p(a_2 x + b_2) := \int_{-\infty}^{\infty} dx\, p(a_1(y - x) + b_1) p(a_2 x + b_2) = p(ay + b)$$

for all (real) constants $a_1>0$, b_1, $a_2>0$, b_2

- The convolution is particularly important for the variable S_N, in fact, its probability distribution, $P(S_N)$, is obtained via successive convolutions of the original distribution p(x) of the x_i variables
- Given $\vec{x} = \{x_1, \cdots, x_N\}$ and the "sum" variable $S_N = \sum_{n=1}^{N} x_n$ formally we have the following probability distribution for S_N:

$$P(z) = \int d\vec{x}\ \delta(z - S_N)\ p(\vec{x}) = \int dx_1 \cdots dx_N\ \delta(z - S_N)\ p(x_1) \cdots p(x_N)$$

- For example let me consider the variable S_2, we have

$$P(z) = \int dx_1\, dx_2\ \delta(z - S_N)\ p(x_1)p(x_2) = \int dx_1\, dx_2\ \delta\left(z - (x_1 + x_2)\right)\ p(x_1)p(x_2) = \cdots$$

$$\cdots = \int dx_1\ p(x_1)p(z - x_1)$$

Which is a convolution of the original pdf, p(x), with itself.

- By iterating on N we would obtain a series of N convolutions

29

30

- The previous equation becomes particularly simple in Fourier space, where the convolution p(y) = f(x)⊗g(x) reduces to a product of the Fourier transforms:

$$p(k) := \mathcal{F}[p(y)] = \mathcal{F}[f(x)]\mathcal{F}[g(x)] = f(k)\, g(k)$$

- Let us consider some examples. Choose a Gaussian as the probability density, and take $a_1=a_2=1$ and $b_1=b_2=0$. Then:

$$\mathrm{F}[p(x) \otimes p(x)] = e^{-\frac{k^2}{2}} e^{-\frac{k^2}{2}} = e^{-k^2}$$

so that a = 1/√2 and b=0.

- Last equation is not specific to the Gaussian; it is also satisfied by exponentials with different arguments. For instance, by the simple exponential

$$e^{-\frac{|k|}{2}} e^{-\frac{|k|}{2}} = e^{-|k|}$$

which is the Fourier transform of the Cauchy distribution:

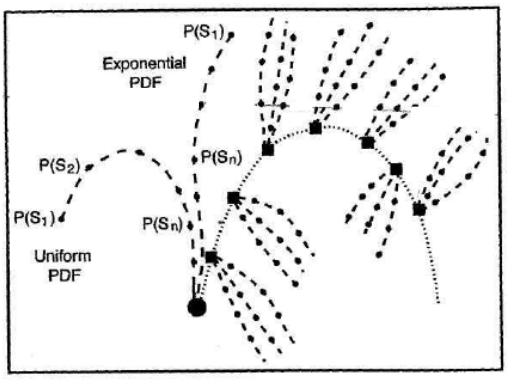
$$\mathcal{F}^{-1}\left[e^{-|k|}\right] = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, e^{ikx} e^{-|k|} = \frac{1}{\pi}\frac{1}{1+x^2}$$

- Therefore, both the Gaussian and Cauchy distributions are, up to scale factors, invariant under convolution and thus stable.

- This means that the variable S_N obtained summing Gaussian (Lorentzian) i.i.d. random variables has a Gaussian (Cauchy-Lorentz) limiting distribution … and the Cauchy-Lorentz case is an example of an extension of the Central Limit theorem.

- Theorem (Levy and Khintchine): A probability density L(x) can only be a limiting distribution of the sum of independent and randomly distributed random variables if it is stable

- Our previous example shows that the Gaussian and Cauchy distributions are limiting distributions. However, there are many more (see supplementary material). In fact, an infinite number of attractors is present in the functional space of pdfs, they comprise the set of all the stable distributions.



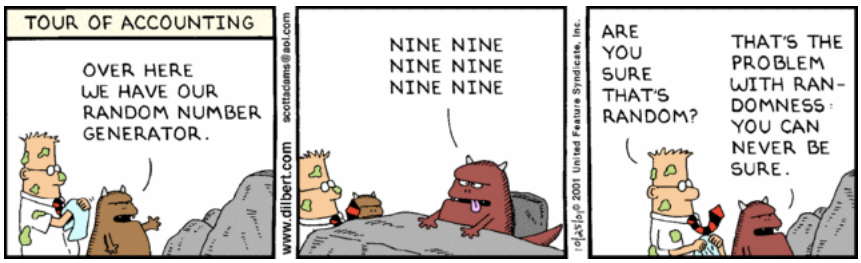
31

Random numbers

- The fundamental constant which links all Monte Carlo methods together, the common factor which makes them all Monte Carlo methods, is that they contain an element of randomness.

- Things happen in a probabilistic fashion in a Monte Carlo calculation, and it is only the average over many unpredictable events that gives us an (approximate) answer to the question we are interested in.

- To generate such unpredictable events, all Monte Carlo methods require a source of random numbers.



Pseudo Random numbers

- The basic problem in random number generation is to generate a sequence of random real numbers r uniformly distributed in the range $0 \le r < 1$.
- If we can do this, then techniques exist for transforming these numbers into random numbers with any other distribution we might be interested in. For example, a random real number R uniformly distributed in some other range $R_{min} < R < R_{max}$ can be derived from r:

$$R = R_{min} + (R_{max} - R_{min})\, r$$

- Another situation which arises frequently in Monte Carlo methods is when we want to perform some action with a certain probability p. In that case, we generate a random real number r as above, and then we perform the action if $r < p$, and not otherwise.
- Almost all the random numbers used in Monte Carlo calculations are in fact only "pseudo-random" numbers, which means that they are generated by a computer program using a deterministic algorithm.
- These pseudo-random number generators can be quite sophisticated and for the most part are adequate for Monte Carlo methods.

- Thus a definition of randomness which solve this paradox in the context of pseudo-random sequences generated by a computer is:
  the deterministic code which produce the random sequence must be statistically uncorrelated with the code which uses it as input
- If two random numbers generators lead to different results when used as source of random numbers for a particular application, then at least one of them is a bad generator for such application

  ⇒ Randomness quality is application dependent

- Desirable properties of random number generators:
  - LONG PERIOD: ideally the generator should give sequences of random numbers that do not repeat themselves
  - UNIFORMITY: the sequence of random number $r \in [0,1)$ should cover uniformly (even in its progression) the interval [0,1)
  - EFFICENCY: the generator should be computationally efficient; one would use as few as possible CPU time in the generation of the sequence
  - UNCORRELATED SEQUENCES: Every sub-sequence should not be correlated to other sub-sequences of the principal one (hard!)
  - REPRODUCIBILITY(!!!): think about you have a subtle bug in your code …

Introduction to pseudo-random number generators

- Most of the common techniques for generating pseudo-random numbers work with integers rather than real numbers.
- Typically they generate an integer i between zero and some maximum value m-1. In order to turn these integers into random real numbers between zero and one, we simply divide by m.
- The simplest and most widely used method of generating a sequence of pseudo-random integers i_n is to act with some function f on the previous value in the sequence:

$$i_{n+1} = f(i_n)$$

- The function f should incorporate only integer arithmetic, such as addition or multiplication or integer division, in order that it always produce integers.
- In order to get the process started we also need to provide the first random number i_o for the sequence. This first number is called the seed for the random number generator.
- It is in fact extremely useful that the generator requires a seed. It means that we can produce different sequences of pseudorandom numbers by feeding different seeds, or we can generate the same sequence more than once by reusing a single seed value.

- This second possibility can come in very handy for debugging Monte Carlo algorithms. If the algorithm shows a bug in a particular run of the program, it may not repeat the alarming behaviour on another run with a different set of random numbers.
- However, we can always duplicate it by using the same seed to reproduce the same set of random numbers, and we can do this as many times as we like until we track down the source of the problem.
- Random number generators of this form will always produce a cycle of integers: if at any point during the sequence we produce a number in which is the same as one we have produced before i_{n-k}, then the previous equation tells us that the sequence will obey $i_n = i_{n-k}$ for all subsequent values of n ⇒the generator will fall into a cycle of length k
- A large part of the task of designing a good random number generator is to find functions f which have a long cycle. More sophisticated generators make use of two or more of the previous values in the sequence:

$$i_{n+1} = f(i_n, i_{n-1}, \cdots)$$

- Generators of this form are also ultimately guaranteed to fall into a repeating cycle of values, for essentially the same reasons as before, although for a generator which uses k previous values of the sequence to generate the next, the cycle can be up to m^k steps long, which could be a very large number indeed for quite modest values of k.

Linear Congruential Generators

- By far the most widely used type of random number generator is the linear congruential generator, first introduced by Lehmer (1951), which is a generator of exactly the type described by $i_{n+1}=f(i_n)$.
- The function used to produce the sequence is:

$$i_{n+1} = (a i_n + c) \bmod m$$

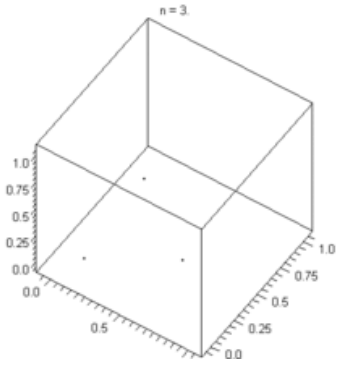
where p mod q represents the modulo operation, which returns the remainder after p is divided by q. (In C it is written as `p%q`, in FORTRAN as `MOD(P,Q)`)

- The highest random number that this equation can produce is m-1, and so the longest sequence of numbers that can be produced has m elements, all of them different.
- Thus in order to make the sequence long we need to make m large. The maximum possible value of m is equal to the largest value that your computer can store in an integer variable. Let us call this number w. On 64 bits computers integers are stored in 64 bits, which means that $w = 2^{64}$

- However, it turns out that the random sequences produced by the previous equation when *m = w* are rather poor in one respect: the low order bits in the binary representation of the numbers will not be very random.
- An extreme example is the lowest order bit, which for m = w will, depending on the values of *a* and *c*, either be a constant or will alternate between 1 and 0 for successive numbers in the series.
- There are other considerations in the choice of values for the constants. The length of the sequence generated is not always equal to *m*; the value of *m* merely provides an upper bound on the length.
- To see this consider for example the case of *a*, *c* and *m* all even. In this case the algorithm will produce only even integers, and clearly the sequence can then be no more than *m*/2 in length. In addition to the length of the sequence there are more subtle problems associated with poor choices for the constants.

Improving the linear congruential generator

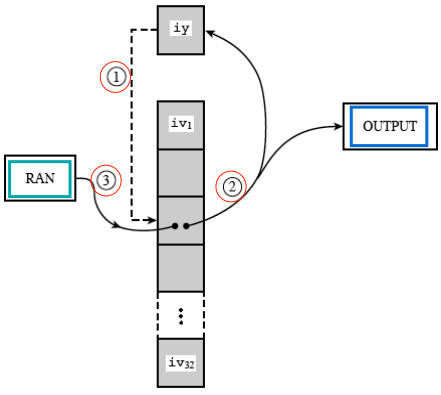
- Although adequate for many applications, linear congruential random number generators have problems. First there is the length of the sequence.
- A subtle problem concerns hidden correlations. It turns out (Knuth, 1981) that if you take the numbers produced by one of these generators in groups of k (which can be any number you like) and regard these groups as the coordinates of points in a k-dimensional space, the points produced are constrained to lie on hyper-planes of co-dimension one in the space. These correlations are certainly well hidden and might not matter for your particular application, but one must be cautious.



- So how can we improve on the linear congruential generator? The most commonly used method is simply to take the random numbers it produces and shuffle them around a bit, to reduce their correlations and increase the repeat period of the generator. This technique requires very little extra computational effort after the numbers have been generated, and much improves the quality of the sequence.

- The standard shuffling scheme, due to Bays and Durham (1976), is this. We create an array of N integers $\{iv_n\}$ and initially we fill it with random integers generated by our linear congruential generator. We also generate one additional random integer, which we store in a variable which we will call iy. Then the procedure goes as follows:

1. Calculate the index $k=\lfloor iy\ N/m \rfloor+1$, which is an integer between 1 and N
2. We return the k^{th} element iv_k from our array as the new random number generated by our shuffling algorithm, and in addition we set $iy = iv_k$
3. We generate a new random integer using our linear congruential generator and we set iv_k equal to this new number



- The function `ran1(…)` from *Numerical Recipes* employs this shuffling scheme on a linear congruential generator, `ran0(…)`

- The numbers generated by this scheme depend on average on the last N numbers generated by the linear congruential generator, and so can be expected to have a repeat cycle on the order of m^N. For our purposes this is essentially infinite and we can stop worrying about the cycle. Furthermore, the subtle correlations concerned with hyper-planes that we mentioned are destroyed by this shuffling of our random numbers, greatly reducing the chances of spurious systematic errors being introduced into our Monte Carlo results.

- Armed with a random number generator, we can now generate random real numbers r_n between 0 and 1 simply by dividing each integer generated by its maximum allowed value plus one:

$$r_n = \frac{i_n}{m} \qquad r_n \in [0,1)$$

and this is the fundamental random quantity which we use in Monte Carlo methods …

we are ready to start!

Sampling random variables

- A sampling procedure is an algorithm that can produce a sequence of values of x ("random variables") x_1, x_2, … such that for any $\Sigma \in \Omega$

$$P(x_k \in \Sigma) = \int_\Sigma p(x)dx \le 1$$

- For a one-dimensional distribution defined on [a,b] this means that

$$P(x_k \in [a,b]) = \int_a^b p(x)dx$$

- It will be possible to do this only by already having a sequence of some basic random variables. It has become conventional to start with random variables that are independent and uniformly distributed on [0,1]. We assume that they can be generated at will by a computer procedure. Such routines have been discussed in the previous lecture.
- Definition: Given a one-dimensional probability distribution p(x) defined on $-\infty<x<+\infty$ the cumulative distribution function F(x) is defined by

$$F(x) = \int_{-\infty}^{x} p(x')dx'$$

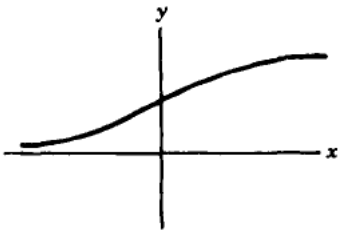
$$\lim_{x\to-\infty} F(x) = 0 \qquad \lim_{x\to\infty} F(x) = 1 \qquad \frac{dF(x)}{dx} = p(x) \ge 0 \qquad P(x_k \in [a,b]) = \int_a^b p(x)dx = F(b) - F(a)$$

Transformation of random variables

- In the discussion that follows, an indefinite supply of uniform pseudo-random variables is assumed to exist.
- Suppose that x is a random variable with cumulative distribution function $F_x(x)$ and probability distribution

$$p_x(x) = \frac{dF_x(x)}{dx}$$

and that y=y(x) is a continuous non-decreasing function of x



- What is $F_y(y)$? And $p_y(y)$?
- The variable x and the function y(x) map in

$$y(x_1) \le y(x_2) \quad iff \quad x_1 \le x_2$$

so the probabilities become

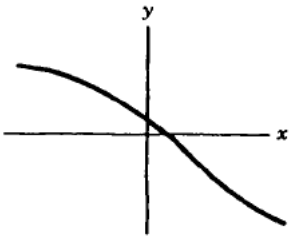
$$P[y(x_1) \le y(x_2)] = P[x_1 \le x_2] \quad or \quad F_y(y) = F_x(x) \quad where \quad y = y(x)$$

- The relationship between the probability distribution functions may be determined by differentiating

$$F_y(y) = F_x(x)$$

This gives

$$p_y(y)\frac{dy}{dx} = p_x(x)$$



- Suppose that *y(x)* is a non-increasing function of x as in the picture; since

$$P[x_1 \ge x_2] + P[x_1 < x_2] = 1$$

then

$$P[y(x_1) \le y(x_2)] = P[x_1 \ge x_2] = 1 - P[x_1 < x_2]$$

- The cumulative distribution for y is therefore

$$F_y(y) = 1 - F_x(x)$$

and

$$p_y(y)\frac{dy}{dx} = -p_x(x)$$

- The probabilities in this equation are non-negative since dy/dx is negative.
- The relationship between the probability distributions of **x** and y for both cases can be combined in one equation as

$$p_y(y) = p_x(x)\left|\frac{dy(x)}{dx}\right|^{-1}$$

Some examples

- Suppose that x is a random variable on [0,1] with $p_x(x) = \frac{4}{\pi}\frac{1}{1+x^2}$ and y=1/x, 1<y<∞ ; then $\frac{dy(x)}{dx} = -\frac{1}{x^2} = -y^2$

  and $$p_y(y) = \frac{4}{\pi}\frac{1}{1+x^2}y^{-2} = \frac{4}{\pi}\frac{1}{1+1/y^2}\frac{1}{y^2} = \frac{4}{\pi}\frac{1}{1+y^2}$$

The probability $p_y(y)$ is a different distribution. In this case, however, it has the same functional form as $p_x(x)$ but on a different range.

- As another example, consider the linear transformation y=ax+b

$$\Rightarrow \quad p_y(y) = |a|^{-1} p_x\left(\tfrac{y-b}{a}\right)$$

- Suppose x is distributed normally with mean 0 and variance 1:

$$p_x(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-x^2/2\right), \quad with \quad -\infty < x < \infty$$

and y is a linear transformation of x, y=σx+μ . Then

$$p_y(y) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right], \quad with \quad -\infty < y < \infty$$

The random variable y is also normally distributed, but its distribution function is centered on μ and has variance σ^2

An important example

- In the discussion so far we have talked about transforming a random variable x, having any distribution, into a random variable y. Because conventional random number generators yield values uniform on [0,1] the transformation from that case is particularly important.
- Given a random variable x with a generic probability distribution $p_x(x)$ and a cumulative distribution $F_x(x)$, consider now $y= F_x(x)$. What is the probability distribution, $p_y(y)$, which characterize y ?

$$p_y(y) = p_x(x)\cdot\left(\frac{dy}{dx}\right)^{-1} = p_x(x)\cdot\left(\frac{dF_x(x)}{dx}\right)^{-1} = p_x(x)\cdot\left(p_x(x)\right)^{-1} = \frac{p_x(x)}{p_x(x)} = 1$$

moreover, being $y= F_x(x)$, it follows that $0 \le y \le 1$

- Thus, y is uniformly distributed on [0,1] , independently from the probability density, $p_x(x)$, which characterizes the random variable x
- If $F_x(x)$ is known and if it is analytically invertible, this has important implications for the statistical sampling of any one-dimensional probability distribution (Ulam, 1947): once sampled $y \in [0,1]$ uniformly (this is our pseudo-random number generator)

$$x = F_x^{-1}(y)$$

is a random variable with probability distribution $p_x(x) = dF_x(x)/dx$

- A first example of this technique is the following: we wish to sample a random variable x with probability distribution given by the Lorentzian function

$$p_x(x) = \frac{1}{\pi}\frac{\Gamma}{\Gamma^2 + x^2}, \quad -\infty < x < +\infty$$

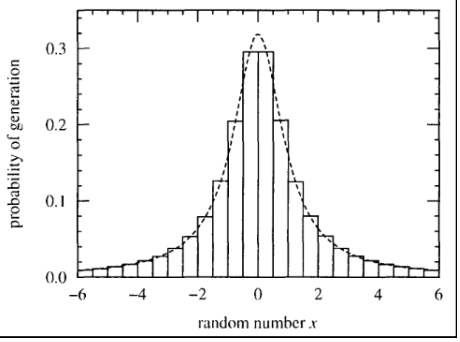
where Γ is the width of the Lorentzian. We have

$$F_x(x) = \int_{-\infty}^{x} \frac{1}{\pi}\frac{\Gamma}{\Gamma^2 + x'^2} dx' = \frac{1}{\pi}\tan^{-1}\frac{x'}{\Gamma}\Big|_{-\infty}^{x} = \frac{1}{\pi}\tan^{-1}\frac{x}{\Gamma} + \frac{1}{2}$$

Let $y=F_x(x) \in [0,1]$ uniformly distributed, then

$$x = \Gamma \tan\left[\pi\left(y - \frac{1}{2}\right)\right]$$

is a random variable with the desired probability distribution

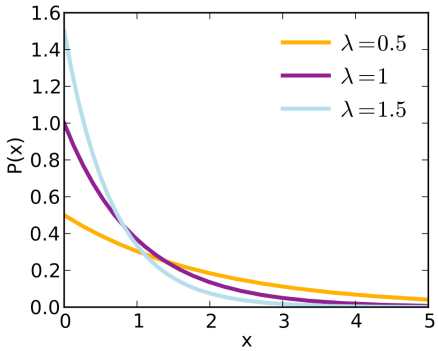


- A second example of this technique is the following: we wish to sample a random variable x with probability distribution given by the exponential function

$$p_x(x) = \lambda \exp(-\lambda x)$$

$$0 \le x < +\infty$$

where λ is the decay rate.



- We have

$$F_x(x) = \int_0^x \lambda e^{-\lambda x'} dx' = -e^{-\lambda x'}\Big|_0^x = 1 - e^{-\lambda x}$$

Let $y=F_x(x) \in [0,1]$ uniformly distributed, then

$$x = -\frac{1}{\lambda}\ln(1 - y)$$

is a random variable with the desired probability distribution. Note that it's safe with $y \in [0,1)$

- Another example: suppose we want to generate a unit vector which points in a random direction in three-dimensional space.
- We can represent a unit vector by the two angles θ and φ of spherical coordinates, such that the components of the unit vector are

$$x = \sin\theta\cos\varphi$$
$$y = \sin\theta\sin\varphi$$
$$z = \cos\theta$$

- We know that the element of solid angle in these coordinates is $\sin\theta d\theta d\varphi$.
- We want to generate random values of θ and φ, such that an equal number of the vectors they describe fall in equal divisions of solid angle. In other words we want to generate uniformly distributed values of φ between 0 and 2π (which is easy) and we want to generate values of θ between 0 and π distributed according to the frequency function

$$p_\theta(\theta) = \frac{1}{2}\sin(\theta)$$

- We have $F_\theta(\theta) = \int_0^\theta \frac{1}{2}\sin\theta' d\theta' = \frac{1}{2} - \cos\theta'\Big|_0^\theta = \frac{1}{2}(1-\cos\theta)$
- Now we can invert this equation to give us θ

$$\theta = \cos^{-1}(1-2r) \quad \textit{with} \quad r \in [0,1] \quad \textit{uniformly distributed}$$

Generating Gaussian random numbers (Box-Muller)

- An issue which arises often is the generation of random numbers which are distributed according to a Gaussian or normal distribution with standard deviation σ:

$$p_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{x^2}{2\sigma^2}\right), \quad \textit{with} \quad -\infty < x < \infty$$

- Gaussian distributed random number x should be generated from a uniform random number generator which produces the number r

$$r = F_x(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{x'^2}{2\sigma^2}\right)dx' = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}\sigma}\right)\right]$$

- where erf(x) is the error function, which is essentially just the definite integral of a Gaussian. Unfortunately, there is no known closed-form expression for the error function, which makes it impossible to invert this equation. Some computer languages, provide a library function which can evaluate erf(x) using an asymptotic series approximation. However, no such library functions exist for evaluating the inverse of the function.

- Generating Gaussian random numbers is extremely important for many applications however, so other methods have been developed to tackle the problem. The standard way of doing it is a two-dimensional variation of the transformation method. Imagine we have two independent random numbers x and y, both drawn from a Gaussian cumulative probability distribution with the same standard deviation σ:

$$p(x,y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$$

We could alternatively express this in polar coordinates as the probability that the point falls in the elemental area r dr dθ with radial coordinate between r and r+dr and angular coordinate between θ and θ +dθ:

$$p(r,\theta) = \frac{r}{2\pi\sigma^2}\exp\left(-\frac{r^2}{2\sigma^2}\right)$$

Thus, if we can generate random values of r and θ according to this distribution and then transform them back into Cartesian coordinates x and y, we will have two random numbers which are distributed according to a Gaussian with standard deviation σ

- Generating the θ variable is trivial, we just need to produce a uniformly distributed real number between zero and 2π.
- The radial coordinate r can be generated using the transformation method. The normalized distribution function for r is

$$p(r,\theta) = \frac{r}{\sigma^2}\exp\left(-\frac{r^2}{2\sigma^2}\right)$$

The transformation method then says that we should produce a value r for this coordinate every time our uniform random number generator produces a number ρ such that

$$\rho = F_r(r) = \int_0^r \frac{r'}{\sigma^2}\exp\left(-\frac{r^2}{2\sigma^2}\right)dr' = 1-\exp\left(-\frac{r^2}{2\sigma^2}\right)$$

rearranging for r, this gives us $r = \sqrt{-2\sigma^2\log(1-\rho)}$

- Note that it's safe with $\rho \in [0,1)$
- With this value for r and our random value for θ, the two numbers

$$x = r\sin\theta, \quad y = r\cos\theta$$

are Gaussian distributed random numbers

- Do not use log(ρ) because sometime ρ=0 with pseudo-random generators
- One can use a linear transformation to centre the Gaussian on μ.

Lecture 1: Suggested books

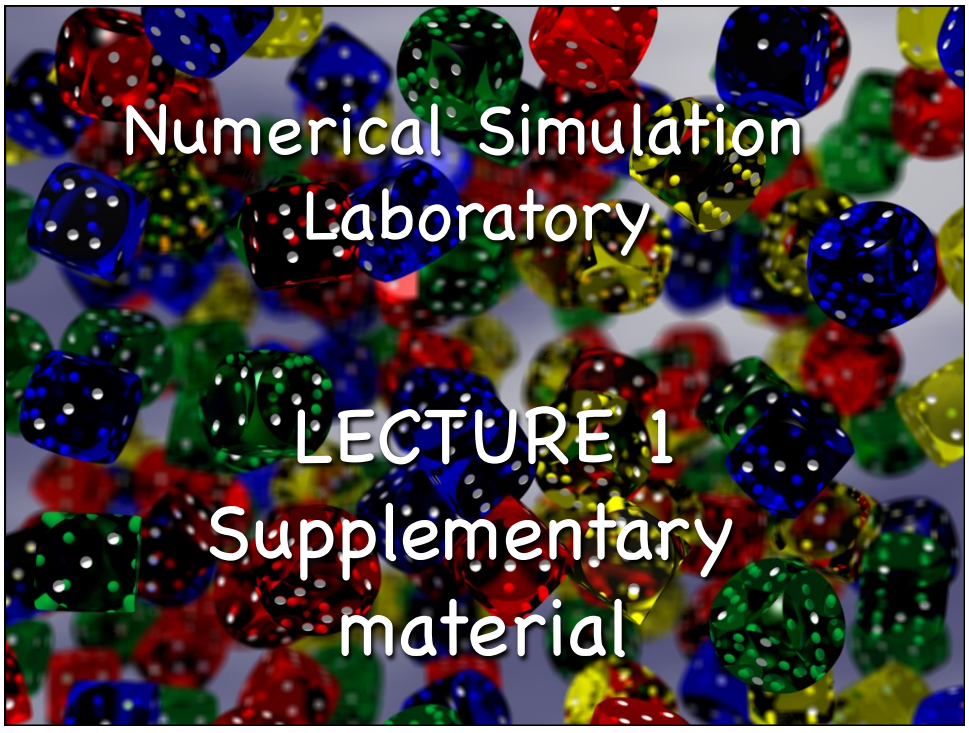
E. Vitali, M. Motta, D.E. Galli, *Theory and Simulation of Random Phenomena*, Springer (2018)

Probability Theory & Central Limit Theorem:

- A. Rotondi, P. Pedroni, A. Pievatolo, *Probabilità, Statistica e Simulazione*, Springer (2012)
- W. Paul & J. Baschnagel, Stochastic Processes, Springer (2013)
- R.N. Mantegna & H. E Stanley, *An Introduction to Econophysics* , Cambridge University press (2000)
- W. Feller, *An Introduction to Probability Theory and its Applications*, Wiley (1966)

Random Numbers Generators & Monte Carlo:

- M.H. Kalos & P.A. Whitlock, Monte Carlo methods, Wiley (2004)
- W. H Press et al., *Numerical Recipes* – Chapter: *Random Numbers*, Cambridge University press (1996)
- Bratley, Fox, Schrage *A Guide to Simulation*, Springer (1987)
- D. Knuth, *The Art of Computer Programming* - Volume 2, Addison-Wesley (1981)



The Characteristic functions

- The characteristic function, $f_X(k)$, corresponding to the stochastic variable, X, is defined as

$$f_X(k) = \left\langle e^{ikx} \right\rangle = \int_{-\infty}^{\infty} dx\, e^{ikx} p(x) = \int_{-\infty}^{\infty} dx \sum_{n=0}^{\infty} \frac{(ikx)^n}{n!} p(x) = \sum_{n=0}^{\infty} \frac{(ik)^n \left\langle x^n \right\rangle}{n!}$$

The previous series expansion is meaningful only if the higher moments, $\langle x^n \rangle$, are small so that the series converges. From this series expansion we see that it requires all the moments to completely determine the probability density, p(x).

- Furthermore, if we know the characteristic function we can obtain moments by differentiating:

$$\left\langle x^n \right\rangle = \lim_{k \to 0} (-i)^n \frac{d^n f_X(k)}{dk^n}$$

this equation provides a simple way to obtain moments if we know $f_X(k)$

- It is relevant to note that the the characteristic function of a Gaussian distribution is

$$f_X(k) = \left\langle e^{ikx} \right\rangle = \int_{-\infty}^{\infty} dx\, e^{ikx} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x - x_0)^2}{2\sigma^2} \right] = \exp\left(ikx_0 - \tfrac{1}{2} k^2 \sigma^2 \right)$$

- Given $\vec{x} = \{x_1, \cdots, x_N\}$ and the variable $A_N = \frac{1}{N} \sum_{n=1}^{N} x_n$ formally we have the following probability distribution for A_N:

$$P(z) = \int d\vec{x}\ \delta(z - A_N)\ p(\vec{x}) = \int dx_1 \cdots dx_N\ \delta(z - A_N)\ p(x_1) \cdots p(x_N)$$

- From this we can calculate the characteristic function of the sum variable

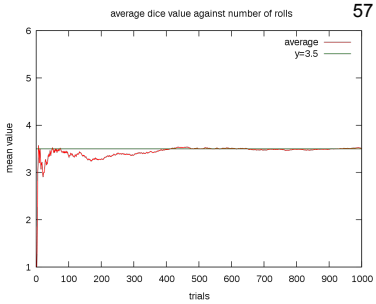
$$\left\langle e^{ikz} \right\rangle = \int_{-\infty}^{\infty} dz\, e^{ikz} \int d\vec{x}\ \delta(z - A_N)\ p(\vec{x}) = \int d\vec{x}\ p(\vec{x}) \int_{-\infty}^{\infty} dz\, e^{ikz}\ \delta(z - A_N) =$$

$$= \int dx_1 \cdots dx_N\ p(x_1) \cdots p(x_N)\, e^{ikA_N} = \left[\int dx_1\ p(x_1)\, e^{ik\frac{x_1}{N}} \right]^N =$$

$$= \left[\exp\left(\frac{ik\mu}{N} - \frac{k^2\sigma^2}{2N^2} + \cdots \right) \right]^N = \exp\left(ik\mu - \frac{k^2\sigma^2}{2N} + o(N^{-1}) \right) \underset{N>>1}{\approx} \exp\left(ik\mu - \frac{k^2\sigma^2}{2N} \right)$$

- Ultimately, the characteristic function therefore approaches the characteristic function of a Gaussian which means that the probability distribution also approaches a Gaussian.

The Law of Large Numbers



- The law of large numbers applies to N independent experiments and may be stated as follows: the average of the results obtained from a large number of trials should be close to the expected value, and will tend to become closer as more trials are performed
- To prove this, the first step involves the derivation of the Chebychev inequality which establishes a relation between the variance and the probability that a stochastic variable can deviate by an amount ε from its average value

$$\sigma_x^2 = \int_{-\infty}^{\infty} dx\,\left(x-\langle x\rangle\right)^2 p(x) \geq \int_{-\infty}^{\langle x\rangle-\varepsilon} dx\,\left(x-\langle x\rangle\right)^2 p(x) + \int_{\langle x\rangle+\varepsilon}^{\infty} dx\,\left(x-\langle x\rangle\right)^2 p(x)$$

$$\geq \varepsilon^2 \int_{-\infty}^{\langle x\rangle-\varepsilon} dx\,p(x) + \varepsilon^2 \int_{\langle x\rangle+\varepsilon}^{\infty} dx\,p(x) = \varepsilon^2 P\left(\left|x-\langle x\rangle\right| \geq \varepsilon\right)$$

where $P(|x-\langle x\rangle|\geq\varepsilon)$ is the probability that the stochastic variable, x, deviates from $\langle x\rangle$ by more than $\pm\varepsilon$.

- We thus obtain the Chebychev inequality

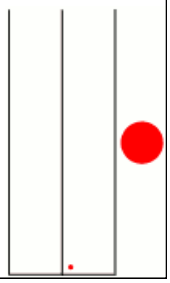
$$P\left(\left|x-\langle x\rangle\right| \geq \varepsilon\right) \leq \frac{\sigma_x^2}{\varepsilon^2}$$

Thus, for fixed variance, the probability that x can differ from its average value by more than $\pm\varepsilon$ decreases as ε^{-2} for increasing ε

- Consider now N independent measurements of the stochastic variable, x. Let S_N be the mean value of the outcomes, $A_N=\sum_n x_n/N$, where x_n , is the outcome of the n-th measurement. Let us first note that $\langle A_N\rangle=\langle x\rangle$. Since we have independent events, the variance behaves as $\sigma_S^2=\sigma_x^2/N$.
- We now use the Chebychev inequality to write

$$P\left(\left|A_N-\langle x\rangle\right| \geq \varepsilon\right) \leq \frac{\sigma_S^2}{\varepsilon^2} = \frac{\sigma_X^2}{N\varepsilon^2} \quad \Rightarrow \quad \lim_{N\to\infty} P\left(\left|A_N-\langle x\rangle\right| \geq \varepsilon\right) = 0 \quad \forall\varepsilon$$

The law of large numbers states that the probability that S_N deviates from $\langle x\rangle$ goes to zero as $N\to\infty$, provided that σ_x is finite.



The Canonical representation of stable distributions

- Levy and Khintchine have completely specified the form of all possible stable distributions
- Theorem (Canonical repesentation): A probability density $L_{\alpha,\beta}(x)$ is stable iff the logarithm of its characteristic function,

$$L_{\alpha,\beta}(k) = \left\langle e^{-ikx} \right\rangle = \int_{-\infty}^{\infty} dx\, e^{-ikx} L_{\alpha,\beta}(k)$$

reads

$$\ln L_{\alpha,\beta}(k) = i\gamma k - c\,|\,k\,|^{\alpha}\left(1 + i\beta\frac{k}{|\,k\,|}\omega(k,\alpha)\right)$$

with $$\omega(k,\alpha) = \begin{cases} \tan\left(\pi\alpha/2\right) & \alpha \neq 1 \\ \left(2/\pi\right)\ln(|\,k\,|) & \alpha = 1 \end{cases}$$

where γ, c, α, and β are real constants taking the values: γ arbitrary, $c \geq 0$, $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, and

60

- The constants γ and c are scale factors. In contrast, α and β define the shape and the properties of $L_{\alpha,\beta}(x)$. These parameters are therefore used as indices to distinguish different stable distributions.
- The parameter α characterizes the large-x; behavior of $L_{\alpha,\beta}(x)$ and determines which moments exist:
  - $0 < \alpha < 2$: Each stable distribution behaves

$$L_{\alpha,\beta}(x) \underset{x\to\pm\infty}{\approx} \frac{1}{|\,x\,|^{1+\alpha}}$$

and has finite absolute moments of order

$$\left\langle |\,x\,|^{\delta} \right\rangle = \int_{-\infty}^{\infty} dx\ |\,x\,|^{\delta}\, L_{\alpha,\beta}(k) < \infty \qquad \text{if } 0 < \delta < \alpha$$

In particular, the latter property implies that the variance does not exist if $\alpha<2$ and that both mean value and variance do not exist if $\alpha<1$

  - $\alpha = 2$: $L_{\alpha,\beta}(x)$ is independent of β, since $\omega(k,\alpha)=0$, and is Gaussian.
- The second characteristic parameter, β, determines the asymmetry of $L_{\alpha,\beta}(x)$:
  - $\beta = 0$: $L_{\alpha,\beta}(x)$ is an even function of x.
  - $\beta = \pm 1$: $L_{\alpha,\beta}(x)$ exhibits a pronounced asymmetry for some choices of α

61

Extension of Central limit theorem

- The previous theorem defines the general expression for all possible stable distributions. However, it does not specify the conditions which the probability density p(x) has to satisfy so that the distribution of the normalized sum S_N converges to a particular $L_{\alpha,\beta}(x)$ in the limit $N\to\infty$
- If this is the case, one can say p(x) belongs to the domain of attraction of $L_{\alpha,\beta}(x)$. This problem has been solved completely:
- Theorem: the probability density p(x) belongs to the domain of attraction of a stable density $L_{\alpha,\beta}(x)$ with characteristic exponent α ($0 < \alpha < 2$, and thus an infinite number!) iff:

$$p(x) \underset{x\to\pm\infty}{\approx} \frac{\alpha a^{\alpha} c_{\pm}}{|x|^{1+\alpha}}$$

where $c_+ \geq 0$, $c_- \geq 0$ and $a>0$ are constants. These constants are directly related to the prefactor c and the asymmetry parameter β by

$$c = \begin{cases} \dfrac{\pi(c_+ + c_-)}{2\alpha\Gamma(\alpha)\sin(\pi\alpha/2)} & \alpha \neq 1 \\ \dfrac{\pi}{2}(c_+ + c_-) & \alpha = 1 \end{cases} \qquad \beta = \begin{cases} \dfrac{c_- - c_+}{c_+ + c_-} & \alpha \neq 1 \\ \dfrac{c_+ - c_-}{c_+ + c_-} & \alpha = 1 \end{cases}$$

62

Bertrand's paradox

- Let us begin with a puzzle:

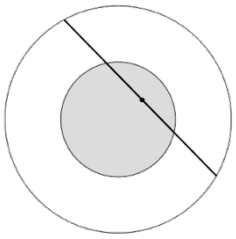
Take a circle of radius 2 in the plane and choose a chord of this circle at random. What is the probability this chord intersects the concentric circle of radius 1 ?

Solution #1

Any such chord is uniquely determined by the location of its midpoint. Thus:

Probability of hitting inner circle =

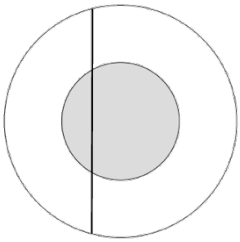
=(area of inner circle)/(area of larger circle)=1/4



Solution #2

By symmetry under rotation we may assume the chord is vertical. The diameter of the large circle is 4 and the chord will hit the small circle if it falls within its diameter (=2). Hence:

Probability of hitting inner circle = 2/4 = 1/2



Bertrand's paradox 2

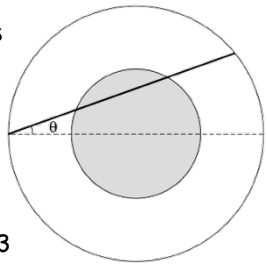
Solution #3

By symmetry we may assume one end of the chord is at the far left point of the larger circle.

The angle θ the chord makes with the horizontal lies between $\pm\pi/2$ and the chord hits the inner circle if θ lies between $\pm\pi/6$.

Therefore:

Probability of hitting inner circle = $(2\pi/6)/(2\pi/2) = 1/3$



- The Bertrand paradox is a problem within the classical interpretation of probability theory.
- Joseph Bertrand introduced it in his work *Calcul des probabilités* (1889) as an example to show that probabilities may not be well defined if the mechanism or method that produces the random variable is not clearly defined.

68

Bertrand's paradox 3

- The problem's classical solution thus depend on the method by which a chord is chosen "at random".
- It turns out that if, and only if, the method of random selection is specified, does the problem have a well-defined solution. There is no unique selection method, so there cannot be a unique solution.
- The three solutions presented correspond to different selection methods, and thus to different random distributions of the chord midpoints, and in the absence of further information there is no reason to prefer one over another
- … unless one is trying to model an effective experiment and thus the word "random" represent a precise way in which one is performing the experiment. In this case only one choice is the right one …
- Anyway, this example shows that we must carefully define what we mean by the term "random" (Note that we use random, aleatory, casual, stochastic as synonyms)