STATISTICAL MECHANICS

Joan J. Cerdà

PART 0 Introduction to statistical mechanics

-**Statistical mechanics: is the tool to link macroscopic physics with microscopic physics (quantum physics).**

-**The aim is to explain the macroscopic properties of the matter from the microscopic properties of atoms and molecules, e.g.**

- **Specific heat.**
- **Transport coefficients: viscosity, etc.**
- **Curie Temperature for a magnet.**
- **State equation for a gas. (P,V,T).**
- **Thermodynamics' laws.**
- **Radial distribution of stars.**

-**The field of statistical mechanics is enormous, e.g. :**

- **Gas in a container, theory of gases.**
- **Behaviour of photons: bulb, laser, etc.**
- **Atoms in a crystal.**
- **Galaxies and stars.**
- **Cars in a high-way.**
- **Chemical reactions, from the most simple to the most complex like BZ.**

-The observed macroscopic physics is a consequence of the microscopic physics when there are many particles in the system.

-In macroscopic systems, the number of particles is enormous:

-22.4litres at NC = $6*10^{23}$ molecules.
-$1\mu m^3$ at NC = $3*10^7$ molecules .

- WE NEED TO USE STATISTICS METHODS TO HANDLE WITH THESE SYSTEMS!!! It is impossible to deal with each one of the particles in the system at an individual level.

- What makes the study of the behavior of these systems non-trivial are the interactions among the particles, and the interaction of the particles with external fields. This can lead for instance to phase transitions, symmetry breaking of the temporal inversion, which cannot be explained as a sum of the parts that form the system.

- *History of the statistical mechanics* -

•Forerunner: kinetic theory of gases → Bernoulli (1738), Herapath (1821), Joule (1851) → Pressure is due to the motion of particles.

• Clausius (1857) → ideal gas law. Concept of mean free path → first to analyze transport phenomena.

• Maxwell (1860) → distribution of molecular speeds. Maxwell's transport equation.

•Boltzmann (1868) → distrib. law for polyatomic gases (aka MB-distribution). Boltzmann factor exp(-βε). Equipartition theorem. H-theorem. How to compute entropy. Transport equation.

- **Ensemble theory** → p,q,,phase point, phase space and trajectories. Mental copies. Density function.

- Maxwell (1879) & Boltzmann (1871) → first to use the idea of ensambles.

- Gibbs (1902) → makes the tool more robust, general and appealing. Relates it to Lagrange's and Hamilton's equations of motion.

- Planck (1900), Einstein (1905), Compton (1923) → quantum leap.

- Bose (1924) → Black body radiation = gas of photons. Photons are indistinguishable particles (aka bosons).

- Einstein (1924) → Bose-Einstein statistics. BEC.

- Fermi (1926) → Some particles cannot occupy the same state. Fermi-Dirac statistics. These kind of parfticles are distinguisable (fermions).

- Sommerfeld (1928) → Theory of metals.

- Reformulation of ensemble theory → need of adapting it to the quantum era.

- Landau, von Neumann & Pauli (1927) → density matrix.

- Belinfante & Pauli (1939) → ah!, the spins determines if particles are bosons or fermions (distinguishable or indistinguishable particles).

- So at the beginnings of the 40’s the statistical mechanics reached its mature age.

- From 40’s to the 2007: many notorious works but mostly concerned with the development or perfection of mathematical techniques. The era of the computers arise and make even more fruitful the research on statistical physics. But all this is fish from another kettle

… to be continue …

STATISTICAL MECHANICS

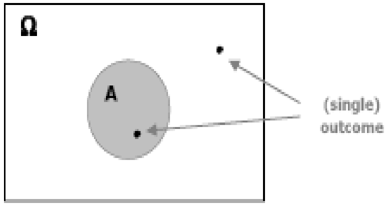
Joan J. Cerdà

PART 1 The theory of probability

Several of the plots in this lesson have been obtained from: http://www.site.uottawa.ca/~nvlajic/ProbabilityTutorial.pdf, by N. Vlajic

Defining probability:

Let us envision an experiment, for which the result is unknown.



Def. 1 **Sample Space (Ω)** – collection of all possible outcomes.

Def. 3 **Probability Space** - three-tuple (Ω, $\mathcal{F}$, Pr), where

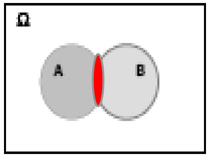
- Ω is a sample space
- $\mathcal{F}$ is a collection of events from the sample space (event space)
- Pr is a probability measure (law) that assigns a number to each event in $\mathcal{F}$

Furthermore, Pr must satisfy, for $\forall$ A, B $\in \mathcal{F}$ the following conditions

(1) $Pr(A) \geq 0$ - probability is a positive measure

(2) $Pr(\Omega) = 1$ - probability is a finite measure

(3) A, B are disjoint events $\Rightarrow$ $Pr(A+B) = Pr(A) + Pr(B)$ - additive property



$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

(1), (2) and (3) are known as axioms of probability measure Pr.

-There are many probabilities, i.e., many functions Pr that satisfy the 3 axioms. Using one of them implies to identify the physical system with the particular behavior implicit in the function Pr . Example tossing a fair or a biased coin.

-One of this functions (for finite sets) is

$$Pr(A) = \frac{cardinal\ of\ A}{cardinal\ of\ \Omega}$$

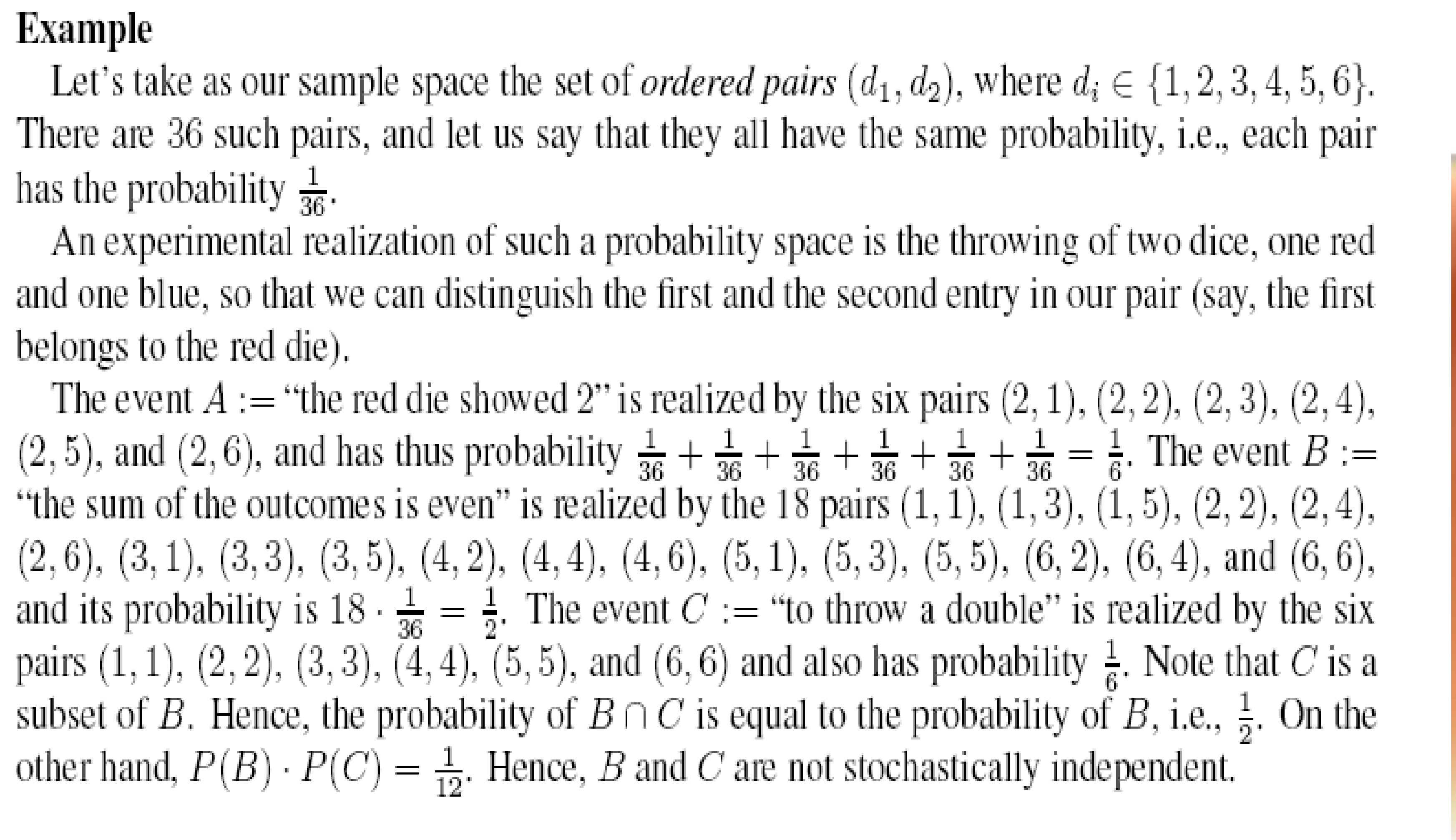
which leads to the 'famous way' of computing probabilities for a given event

'Relative Frequency' Definition of Probability

Perform an experiment a number of times/trials (n), counting the occurrences of event A (n_A). Then the probability P(A) of event A can be found/defined as the limit:

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n}$$

but probabilities are much more than the previous formula !!! …



Example

Let's take as our sample space the set of *ordered pairs* (d_1, d_2), where $d_i \in \{1, 2, 3, 4, 5, 6\}$. There are 36 such pairs, and let us say that they all have the same probability, i.e., each pair has the probability $\frac{1}{36}$.

An experimental realization of such a probability space is the throwing of two dice, one red and one blue, so that we can distinguish the first and the second entry in our pair (say, the first belongs to the red die).

The event $A :=$ "the red die showed 2" is realized by the six pairs $(2, 1)$, $(2, 2)$, $(2, 3)$, $(2, 4)$, $(2, 5)$, and $(2, 6)$, and has thus probability $\frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{1}{6}$. The event $B :=$ "the sum of the outcomes is even" is realized by the 18 pairs $(1, 1)$, $(1, 3)$, $(1, 5)$, $(2, 2)$, $(2, 4)$, $(2, 6)$, $(3, 1)$, $(3, 3)$, $(3, 5)$, $(4, 2)$, $(4, 4)$, $(4, 6)$, $(5, 1)$, $(5, 3)$, $(5, 5)$, $(6, 2)$, $(6, 4)$, and $(6, 6)$, and its probability is $18 \cdot \frac{1}{36} = \frac{1}{2}$. The event $C :=$ "to throw a double" is realized by the six pairs $(1, 1)$, $(2, 2)$, $(3, 3)$, $(4, 4)$, $(5, 5)$, and $(6, 6)$ and also has probability $\frac{1}{6}$. Note that C is a subset of B. Hence, the probability of $B \cap C$ is equal to the probability of B, i.e., $\frac{1}{2}$. On the other hand, $P(B) \cdot P(C) = \frac{1}{12}$. Hence, B and C are not stochastically independent.

Def. 4 **Conditional Probability** enables us to determine whether two events, A and B, are related in the sense that knowledge about the occurrence of one alters the likelihood of occurrence of the other.

probability of A given B has occurred:

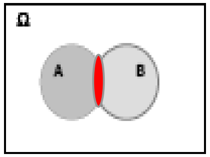
$$\Pr(A|B)=\frac{\Pr(AB)}{\Pr(B)}$$

A occurs, in reduced event space, only if A∩B occurs

reduced event space

consequently:

$$\Pr(AB)=\Pr(A|B)\cdot\Pr(B)$$



$$\Pr(AB)\equiv\Pr(A\cap B)$$

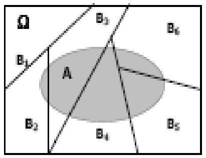
Def. 5 **Independent Events** - two events, A and B, are independent if

$$\Pr(AB)=\Pr(A)\cdot\Pr(B)$$

From Def. 4 and Def. ⇒ A and B are independent if Pr(A|B)=Pr(A).

Theorem 1 **Total Probability**

Let B_1, .., B_n be mutually exclusive events whose union equals the sample space Ω.



Then, the probability of any given event A⊆Ω can be expressed as

$$\Pr(A)=\Pr(A|B_1)\cdot\Pr(B_1)+\Pr(A|B_2)\cdot\Pr(B_2)+..+\Pr(A|B_n)\cdot\Pr(B_n)$$

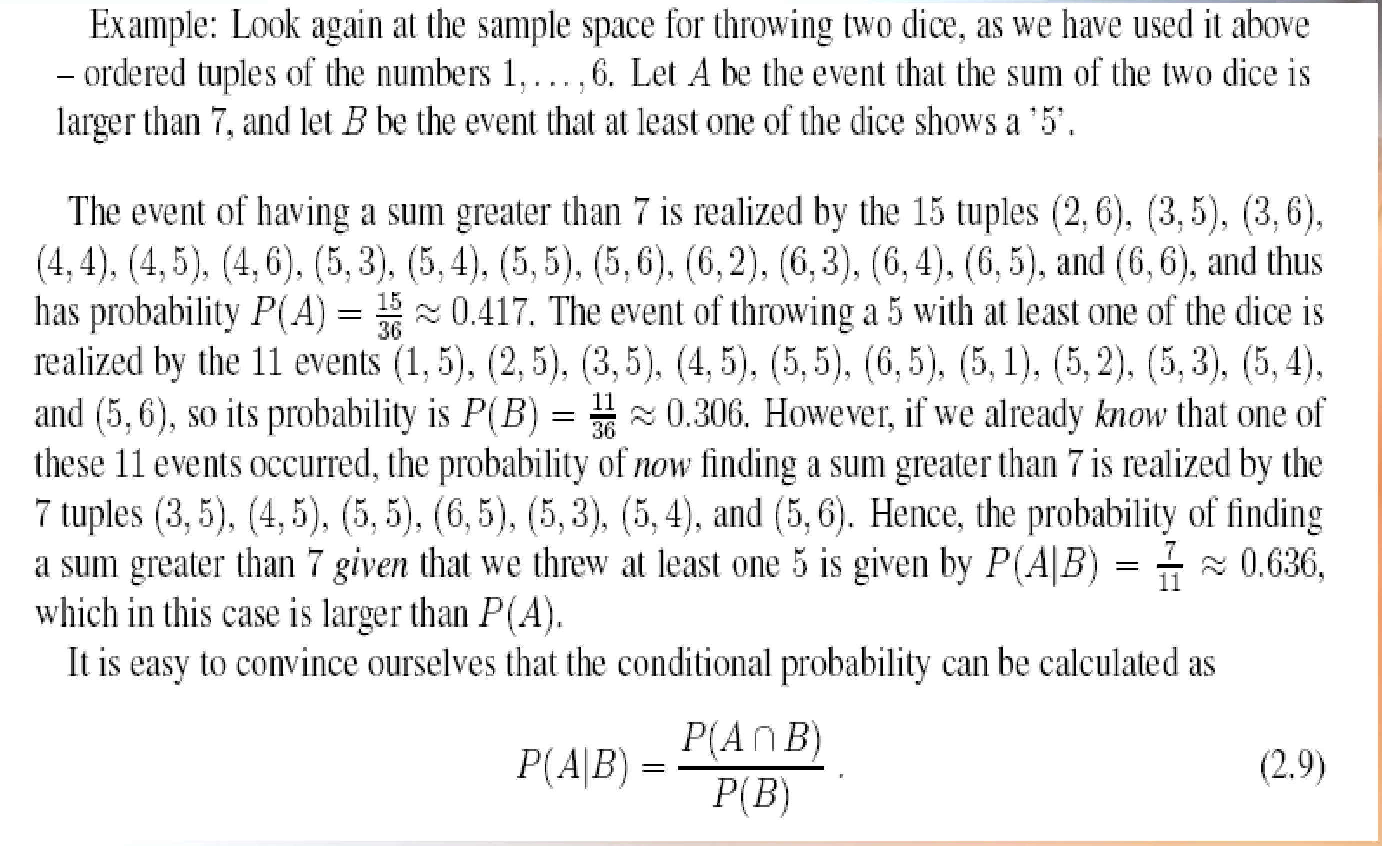
Proof: based on $A = A\cap\Omega = A(B_1+B_2+..+B_n) = AB_1 + AB_2 + .. + AB_n$

- Bayes' Theorem:

$$P(B_j|A)=\frac{P(A|B_j)\,P(B_j)}{\sum_{i=1}^{N} P(A|B_i)\,P(B_i)}$$

-Why is Bayes' theorem important? → suppose that we have two events, where A is the consequence, and B is the cause which in fact it stands for several different possible reasons labeled B_1,B_2, …B_i,…B_N.

- It can happen that it is easy to know $P(A|B_i)$ but we would like to know the reverse $P(B_j|A)$. Bayer's theorem is then very helpful, and does the same work that doing a tree diagram which can be very tedious if the number of B causes is large.



Example: Look again at the sample space for throwing two dice, as we have used it above – ordered tuples of the numbers $1, \ldots, 6$. Let A be the event that the sum of the two dice is larger than 7, and let B be the event that at least one of the dice shows a '5'.

The event of having a sum greater than 7 is realized by the 15 tuples $(2,6)$, $(3,5)$, $(3,6)$, $(4,4)$, $(4,5)$, $(4,6)$, $(5,3)$, $(5,4)$, $(5,5)$, $(5,6)$, $(6,2)$, $(6,3)$, $(6,4)$, $(6,5)$, and $(6,6)$, and thus has probability $P(A) = \frac{15}{36} \approx 0.417$. The event of throwing a 5 with at least one of the dice is realized by the 11 events $(1,5)$, $(2,5)$, $(3,5)$, $(4,5)$, $(5,5)$, $(6,5)$, $(5,1)$, $(5,2)$, $(5,3)$, $(5,4)$, and $(5,6)$, so its probability is $P(B) = \frac{11}{36} \approx 0.306$. However, if we already *know* that one of these 11 events occurred, the probability of *now* finding a sum greater than 7 is realized by the 7 tuples $(3,5)$, $(4,5)$, $(5,5)$, $(6,5)$, $(5,3)$, $(5,4)$, and $(5,6)$. Hence, the probability of finding a sum greater than 7 *given* that we threw at least one 5 is given by $P(A|B) = \frac{7}{11} \approx 0.636$, which in this case is larger than $P(A)$.

It is easy to convince ourselves that the conditional probability can be calculated as

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \; . \tag{2.9}$$

-Binomial probability (or Bernoulli's experiment):

* We have an experiment in which can happen A or Ac.\
* A has associated a probability p.
* Ac has associated a probability q=1-p.
* Repeating the experiment does not change neither p nor q.

we wonder: which is the probability that after doing N of such experiments, In k of them the event A has happened ?

Bernoulli solved that problem long time ago:

$$P_N(k)=\binom{N}{k} p^k q^{N-k}$$

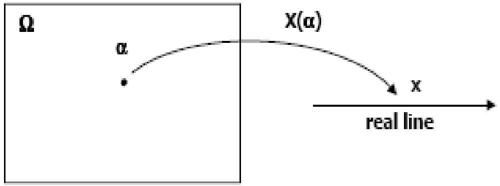
Combinatorial numbers

$$\binom{N}{k}=\frac{N!}{k!\,(N-k)!}$$

Striling's approach

$$n!\sim\sqrt{2\pi n}\, n^n e^{-n}$$

Def. 6 **Random Variable** (X) is a function that assigns a real number (X(α)) to each outcome α in the sample space.



Def. 7 **Continuous Random Variable** - takes on an uncountablely infinite number of distinct values.

Def. 8 **Discrete Random Variable** - takes on a finite or countably infinite number of distinct values.

Def. 9 **Cumulative Distribution Function (cdf) $F_X(x)$** of a random variable X is defined as the probability of the event {X≤x}.

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

consequently:

$$\Pr[a < X \le b] = F_X(b) - F_X(a)$$

$$\Pr[X > x] = 1 - F_X(x)$$

Def. 10 **Probability Density Function (pdf) $f_X(x)$**, if it exists, is defined as a derivative of $F_X(x)$.

$$f_X(x) = \frac{dF_X(x)}{dx}$$

consequently:

$$\Pr[X \le a] = F_X(a) = \int_{-\infty}^{a} f_X(x)dx$$

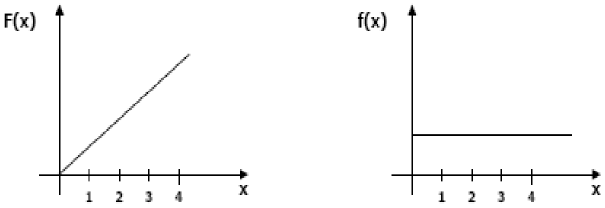
Note: $f_X(x)$ is called "density of probability" at point x, since the probability that X is in a small interval in the vicinity of x is approximately $f_X(x) \cdot x$.

Example: Take the sample space as the set of *ordered pairs* (d_1, d_2), where $d_i \in \{1,2,3,4,5,6\}$ (e.g., throwing of two dice of different color). We could define the random variable S as the function which returns the sum of the numbers which the dice show. Hence, the pair $(2,3)$ will be mapped by S to the value 5.

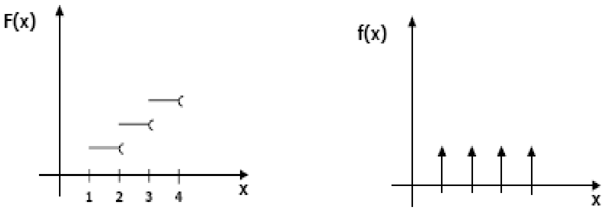
Given some random variable X, we can use it to define particular events related to it. Most useful is going to be the following consideration: Given some specific value x, which elementary events will be mapped by the random variable onto this particular value x?

Example: Take the random variable S defined above. Which events are mapped by S onto the value 4? Answer: $(1,3)$, $(2,2)$, and $(3,1)$. Hence, since there are 36 elementary events, the probability of finding an event which is mapped onto the value 4 is $3/36 = 1/12$. In more colloquial terms: The probability of getting the sum 4 when throwing two dice is $1/12$!

example: **cdf & pdf of a continuous r.v.**



example: **cdf & pmf (prob. mass func.) of a discrete r.v.**



We are often concerned with some characteristic of a random variable rather than the entire distribution:

Def. 11 mean of continuous r.v. $E[X]=\mu_x=\int_{-\infty}^{\infty} x\cdot f_x(x)dx$

Def. 12 mean of discrete r.v. $E[X]=\sum_{\text{all k}} k\cdot \Pr(x=k)$

important properties: $E[a\cdot X]=a\cdot E[X]$ and $E[X+Y]=E[X]+E[Y]$

Def. 13 variance (dispersion around mean) $Var[X]=E[(X-\mu_x)^2]=E[X^2]-\mu_x{}^2$

Def. 14 standard deviation $\sigma_x=\sqrt{Var[X]}$

important properties: $Var[a\cdot X]=a^2\cdot Var[X]$

Example: For the random variable S defined above, the probability distribution $w_S(x)$ looks like this. It is zero everywhere, except for the case when $x\in\{2,3,4,5,6,7,8,9,10,11,12\}$.

$$w_S:\begin{cases} 2\longrightarrow 1/36\approx 0.0278\\ 3\longrightarrow 2/36\approx 0.0556\\ 4\longrightarrow 3/36\approx 0.0833\\ 5\longrightarrow 4/36\approx 0.1111\\ 6\longrightarrow 5/36\approx 0.1389\\ 7\longrightarrow 6/36\approx 0.1667\\ 8\longrightarrow 5/36\approx 0.1389\\ 9\longrightarrow 4/36\approx 0.1111\\ 10\longrightarrow 3/36\approx 0.0833\\ 11\longrightarrow 2/36\approx 0.0556\\ 12\longrightarrow 1/36\approx 0.0278 \end{cases}$$

Note that this probability distribution is *normalized*:

$$\sum_{n=2}^{12} w_S(n) \;=\; 1\,.$$

$$\begin{aligned}\langle S\rangle &= \sum_{n=2}^{12} n\, w_S(n)\\ &= 2\frac{1}{36}+3\frac{2}{36}+4\frac{3}{36}+5\frac{4}{36}+6\frac{5}{36}+7\frac{6}{36}+8\frac{5}{36}+9\frac{4}{36}+10\frac{3}{36}+11\frac{2}{36}+12\frac{1}{36}\\ &= 7\,.\end{aligned}$$

Second example: A friend suggests the following "game": He will flip a coin and we count, how often we get successive "heads" (starting with the first throw!). If we get n heads, he will pay us n Euro. He demands that we pay a price of 2 Euro to him in order to play the game. We could get tremendously rich that way, for just 1 Euro! Should we do this?

Let's first say that N is the random variable which counts how often we had "head" in a row, starting from the first throw. What is the probability distribution of N? Evidently, the probability for getting no head at all ($N=0$, the series right away starts with a tail) is $1/2$. So with probability $1/2$ we don't get anything. We will get exactly 1 head if the first throw is a head and the second is a tail, so the probability is $1/4$. Hence, the probability for exactly n heads in a row at the beginning is

$$w_N(n) \;=\; \frac{1}{2^{n+1}}\,. \tag{2.20}$$

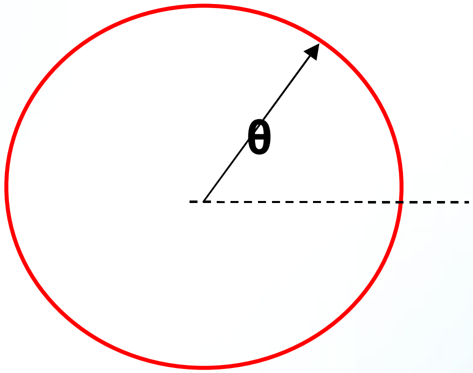
This is normalized:

$$\sum_{n=0}^{\infty} w_N(n)=\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\cdots \;=\; 1\,. \tag{2.21}$$

And the expectation value therefore is

$$\begin{aligned}\langle N\rangle &= \sum_{n=0}^{\infty} n\, w_N(n)\\ &= \sum_{n=0}^{\infty} n\, 2^{-(n+1)}\\ &= 0\frac{1}{2}+1\frac{1}{4}+2\frac{1}{8}+3\frac{1}{16}+4\frac{1}{32}+\cdots\\ &\;\vdots\\ &= 1\end{aligned} \tag{2.22}$$

Example: direction of a vector in a 2D space



$$\theta \in [0, 2\pi]$$

$$f_\theta(\theta) = \frac{\text{١}}{\text{٢}\pi}$$

Let’s suppose we have another random variable x=cos$^{2(}$θ), and we want to compute the expected value of x, i.e. <x> = E(x)

$$x>=\int_0^{\text{٢}\pi} x(\theta)\, f_\theta(\theta)\, d\theta = \int_0^{2\pi} \cos^2(\theta)\, \frac{1}{2\pi}\, d\theta = \pi$$