Probabilities

Examples of stochastic processes:
(i) Tossing a coin,
(ii) Throwing dice,
(iii) Drawing cards from a pack,
(iv) Randomly selecting a point (on a line, plane, etc.).

Examples of stochastic events:
(a) Get #1 when throwing a die,
(b) Get the total of 5 when throwing two dice,
(c) Get some even number when throwing a die,
(d) Get heads when tossing a coin,
(e) Get the arrangement (heads, heads, tails) when tossing a coin three times,
(f) Get two cases of heads and one case of tails when tossing a coin three times,
(g) Draw the Queen of hearts,
(h) Draw a Queen,
(i) Get something of hearts,
(j) Get \( x = \pi \) when randomly selecting a number \( x \) within the interval \( [0, 10] \),
(k) Get \( 0 \leq x \leq 0.5 \) when randomly selecting a number \( x \) within the interval \( [0, 1] \).

Algebra of events
Some events can be decomposed into a combination of more 'elementary' events. Such events are called compound events. In the above list the compound events are (b), (c), (f), (h), (i), and (k). For example, in the case (b), 'total of 5' = '(1,4) or (2,3) or (3,2) or (4,1)'. In the case (f), 'two heads and one tails' = '(heads, heads, tails) or (heads, tails, heads) or (tails, heads, heads)'. Events that cannot be decomposed into simpler ones are called simple events. In our list, the simple events are (a), (d), (e), (g), and (j).

When decomposing compound events, we have used logical operation 'OR'. We can also employ logical operations 'AND' and 'NOT'. Example 1: 'hearts' AND 'Queen' = 'Queen of hearts'. Note that with the operation 'AND' we have produced a simple event out of two compound ones. Example 2: NOT 'heads' = 'tails'. Conclusion: The algebra of events is the same as the algebra of logic.

The algebra of logic is known to be equivalent to the algebra of sets. Hence, one can expect some important correspondence between sets and events. And this is indeed the case. Let us view simple events as some points in the so-called sample space. Then compound events are associated with the sets containing more than one element, the single-element sets corresponding to the simple events. The set representation of events proves especially convenient for calculation of probabilities (see below).

Given the above correspondence with logic and sets, the following notation is used: 'OR' \( \equiv \bigcup \equiv '+' \); 'AND' \( \equiv \bigcap \equiv ' \cdot ' \). Hence, the identical expressions \( A \ OR \ B \equiv A \cup B \equiv A + B \) denote the event when either \( A \), or \( B \), or both happen. The identical expressions \( A \ AND \ B \equiv A \cap B \equiv A \cdot B \) denote the event when \( A \) and \( B \) happen simultaneously. There is a shorthand notation for the operation 'NOT': \( \bar{A} \equiv \text{NOT } A \).

Mutually exclusive events. The events are mutually exclusive if they cannot occur simultaneously. For theoretical purposes, it is convenient to introduce the notion of improbable event, \( \emptyset \),—the event which never happens, the analog of the empty set, and the identically false statement. Then, two
events, $A$ and $B$, are mutually exclusive if and only if $A \cap B = \emptyset$.

**Statistical probability**

Suppose one repeats/observes some stochastic process and counts the number of outcomes when the event $A$ happens. Very often (but not always!) this number, $N_A$, behaves in such a way that if the number of trials, $N$, is large enough, then the ratio $N_A/N$ is practically independent of $N$, approaching in the limit $N \to \infty$ some value $P_A$. In this case we say that the event $A$ has the probability

$$P_A = \lim_{N \to \infty} \frac{N_A}{N}. \quad (1)$$

By its definition, $0 \leq P_A \leq 1$. It is also clear that

$$P_A = 1 - P\bar{A}, \quad (2)$$

since $N_A + N\bar{A} \equiv N$.

**Classical probability**

In some cases, both the existence of a probability and its value are immediately clear from the symmetry of the problem. For example, we have little doubt that the probability of the event (a) is 1/6, the probability of the event (d) is 1/2, the probability of the event (g) is 1/52 (if there are different 52 cards in our pack and the Queen of hearts is one of them)—Why?—Because we see that in each of these cases there is a finite number, $m$, of absolutely equivalent simple events, and the event of interest is one of them. Hence, on the basis of this symmetry, we conclude that if $N$ is large enough, then $N_A \approx N/m$, so that $p = 1/m$. Such cases are called classical probabilities.

**Probability of a sum of mutually exclusive events**

If the events $A$ and $B$ are mutually exclusive, then that $N_{A+B} = N_A + N_B$ (the absence of overlap guarantees the absence of double counting), and from (1) we have

$$P_{A+B} = P_A + P_B. \quad (3)$$

We are interested in predicting the statistical probability of this or that event. In a lot of cases, the idea of classical probability in combination with Eq. (3) is a powerful tool for doing that. We decompose some compound event $A$ into a sum of mutually exclusive simple events—note that by the very definition, any two simple events are either mutually exclusive, or identical—and then count the number $m_A$ of simple events forming the event $A$. Then, $P_A = m_A P_{\text{smpl}} = m_A/m$, where $P_{\text{smpl}} = 1/m$ is the probability of a simple event, with $m$ the total number of all possible simple events. For example, the probability of the event (b) equals 1/9 ($m_A = 4$ and $m = 36$).

**Problem 1.** I select at random a real number $x$ within the interval $[0,1]$—each number having equal chance to be selected. I consider my selection 'successful' if $x < 0.2$ or $x > 0.7$. What is the probability of a successful selection?

For more instructive examples, see “Number $\pi$ by Monte Carlo”.

**Combinatorial analysis: Arrangements, Permutations, Combinations, etc.**

Calculating the number of simple events forming a given compound event is not always trivial and, in a general case, involves combinatorial analysis. Below we mention some typical situations.

*Arrangements of $n$ dissimilar objects.* Example: There are 6 ways of arranging 3 different objects:
ABC, ACB, BAC, BCA, CAB, CBA. In general, the number $A_n$ of all the arrangements of $n$ dissimilar objects equals

$$A_n = n!.$$ (4)

To arrive at the result (4) one can first get the relation $A_n = nA_{n-1}$ by noting that all the arrangements of $n$ elements can be viewed as putting some element on the 1st place ($n$ ways of doing that because of $n$ different elements) and arranging the rest of the $(n-1)$ elements with respect to the other $(n-1)$ places ($A_{n-1}$ different ways, by definition). Then Eq. (4) is clear by induction.

*Multiple-counting trick.* What if there are identical objects in our set of $n$ objects. [Example: AAB, ABA, BAA.] There is a generic trick which is useful in many situations like that. First we ignore the fact that there are identical objects, by writing, say, $A'A''$ instead of AA. This reduces the problem to the previous one—that of arranging dissimilar objects. Then we take into account the multiple counting of one and the same situation. Normally, this is easy, because multiple counting is associated with some factor. If among our $n$ objects we have one (and only one) group of $p$ identical objects, then the multiple counting comes from all possible arrangements of these objects, that is the extra factor is $p!$, and the proper answer is $n!/p!$. If there are more than one group of identical objects, the factor $p_1!p_2!\ldots$ arises for each $i$-th group, and the answer is $n!/p_1!p_2!\ldots$ .

**Problem 2.** I form a string of letters from my first and last names: 'borissvistunov’. Then I select at random, letter by letter, the symbols of this string to form a new string. What is the probability that I get the string 'svistunovboris’?

*Permutations of $m$ objects from $n$.* Now we chose $m$ objects out of $n$ and then arrange them. The number of all different ways of doing that equals

$$P^m_n = \frac{n!}{(n-m)!}.$$ (5)

This result can be obtained by the multiple-counting trick. We represent all the arrangements of $n$ objects as strings of $n$ symbols and declare that the first $m$ positions in the string represent our permutations. The multiple counting comes from the next $(n-m)$ positions. There are $(n-m)!$ different arrangements of symbols in these positions. Hence, we get the extra factor of $(n-m)!$ and need to divide by it.

**Problem 3.** You draw 3 cards from a 52-card pack. What is the probability that you get Queen-King-Ace (the order is important, the suits are not)?

*Combinations of $m$ objects from $n$.* Once again $m$ objects out of $n$, but now we do not care about the order of the elements. That is we are dealing with $m$-element sets rather than strings. Comparing this to permutations (once again multiple-counting trick), we see that we have an extra factor of $m!$ corresponding to arranging $m$ objects to produce permutations from combinations. For the number of combinations we thus get

$$C^m_n = \frac{n!}{m!(n-m)!}.$$ (6)

**Problem 4.** You draw 3 cards from a 52-card pack. What is the probability that all the three are aces?

**Problem 5.** (An example of the so-called *binomial distribution.*) You toss $n$ coins (or one coin $n$ times). What is the probability of getting exactly $m$ heads?
Probability of a product of independent events

Suppose we know the probability $P_A$ of the event $A$ and the probability $P_B$ of the event $B$. Is it enough to establish the probability of the event $A \cdot B$? In a general case—not. Indeed, if $A = B$, then $P_{A,B} = P_A = P_B$. If $A$ and $B$ are mutually exclusive, then $A \cdot B = \emptyset$ and $P_{A,B} = P_{\emptyset} = 0$. There is, however, a typical case where we can readily find $P_{A,B}$—the case when $A$ and $B$ are independent. In accordance with (1),

$$P_{A,B} = \lim_{N \to \infty} \frac{N_{A,B}}{N}.$$  \hfill (7)

If $A$ and $B$ are independent, then in the limit of $N \to \infty$ we have $N_{A,B} \to N_A P_B$. Indeed, the events when $A$ occurs can be considered as independent trials for the event $B$.—Basically, this is what one means by the words 'B is independent of $A$'. So, we expect $N_A P_B$ events $B$ out of $N_A$ trials. [The other trials are not interesting for us, since at these trials $A$ has not take place.] Plugging $N_{A,B} = N_A P_B$ into (7), we find

$$P_{A,B} = P_A P_B. \hfill (8)$$

Suppose that for some two events $A$ and $B$ the experiment yields the relation (8).—Are these events independent? For all practical purposes—yes. Because there is no other way to observe correlations between the two events.

Problem 6. The product of the events (h) and (i) is the event (g)—see the list on the first page. The events (h) and (i) are independent: The fact that a card is a Queen tells absolutely nothing about its possible suit; and vice versa—a particular suit of a card does not change the chances for this card to be a Queen. Correspondingly, we should have $P_{(g)} = P_{(h)} P_{(i)}$. Make sure that this is the case by independently calculating $P_{(g)}$, $P_{(h)}$, and $P_{(i)}$ as classical probabilities.

The $A \to \bar{A}$ trick. What is the probability of getting at least one 'one' when throwing a die $m$ times (equivalently, throwing $m$ dice)? First we use a trick of considering the event $\bar{A}$ instead of $A$. Indeed, we see that in our case $\bar{A} = A_1 \cdot A_2 \cdot A_3 \cdot \ldots \cdot A_m$, where $A_j$ is the event of not having 'one' at the $j$-th attempt (on $j$-th die). [Clearly, $P_{\bar{A}_j}$ is independent of $j$ and is equal to 5/6.] Hence, we can iteratively apply Eq. (8) to get $P_{A} = (5/6)^m$, and $P_{A} = 1 - P_{A} = 1 - (5/6)^m$.

Conditional Probabilities

The probability for the event $B$ to occur under the condition that the event $A$ takes place is called conditional probability; we will denote it as $P_{B/A}$. From the definition (1) it follows that

$$P_{B/A} = \lim_{N \to \infty} \frac{N_{A,B}}{N}.$$  \hfill (9)

Indeed, now the number $N_A$ has the meaning of the total number of trials, since we are interested in the event $B$ only when $A$ does take place. There is a very useful relation between $P_{B/A}$, $P_{A,B}$, and $P_A$. We have

$$P_{A,B} = \lim_{N \to \infty} \frac{N_{A,B}}{N} \equiv \lim_{N \to \infty} \frac{N_A N_{A,B}}{N_A},$$  \hfill (10)

and from (9) and (1) conclude that

$$P_{A,B} = P_A P_{B/A}. \hfill (11)$$

Eq. (11) generalizes Eq. (8), since if $B$ is independent of $A$, then $P_{B/A} \equiv P_B$ and (11) yields (8). The utility of (11) is clear from the following example. What is the probability that when drawing three cards from a 52-card pack we get the Queen of hearts? We start with the $A \to \bar{A}$ trick. We have $\bar{A} = \bar{A}_1 \cdot \bar{A}_2 \cdot \bar{A}_3$, where $A_j$ is the event when the $j$-th card is not the Queen of hearts. Next we iteratively use Eq. (11): $P_{\bar{A}_1 \cdot \bar{A}_2} = P_{\bar{A}_1} P_{\bar{A}_2} \bar{A}_1 = (51/52) \cdot (50/51)$; $P_{A_1 \cdot A_2 \cdot A_3} = P_{A_1 \cdot A_2} P_{A_3} \bar{A}_1 \cdot A_2 = \ldots$
\[ P_{A_1 \cdot A_2} \cdot \frac{49}{50} = \left( \frac{51}{52} \cdot \frac{50}{51} \right) \cdot \frac{49}{50} = \frac{49}{52}. \] Hence, \( P_A = 1 - P_{\bar{A}} = 1 - \frac{49}{52} = \frac{3}{52} \approx 0.0577 \) — not a big chance.

One may notice that for the above example there is an alternative way to find the probability \( P_A \) — by employing summation of mutually exclusive events: \( A = B_1 + B_2 + B_3 \), where \( B_j, j = 1, 2, 3 \) is the event when exactly the \( j \)-th card is the Queen of hearts. Obviously, \( P_{B_j} = \frac{1}{52} \) and one immediately gets the answer. Nevertheless, the way we solved the problem is very instructive, because it gives a clue to solving more difficult problems, where one cannot avoid using conditional probabilities.

**Problem 7.** This year we have some 25 people in our class. What is the probability that at least two of us have the same birthday? For simplicity, assume that there are 365 days in a year.

**Problem 8.** Solve Problem 3 by conditional probabilities. (If you were solving it this way previously, then now solve it by permutations!)

**Radioactive decay**

A very good physical example of a conditional probability is the decay of a radioactive nucleus. Suppose we start observing a radioactive nucleus at the time moment \( t = 0 \). What is the probability, \( P_{\text{surv}}(t) \), that the nucleus survives up to the time \( t \)? The most striking fact about radioactive nuclei is the absence of aging. If a nucleus survives by the time moment \( t \), it is as “fresh” as it was at \( t = 0 \), in the sense that its probability to decay after an arbitrarily long survival remains the same. If for our nucleus we introduce a conditional probability \( P_{\text{decay}}(\Delta t/t) \) to decay during some small time interval \( \Delta t \) after remaining intact during the time \( t \), then the mathematical expression for the above-mentioned physical fact will be

\[
P_{\text{decay}}(\Delta t/t) = \gamma \Delta t \quad (\Delta t \ll \gamma^{-1}),
\]

where \( \gamma \) is a time-independent constant characterizing the decay rate (\( \gamma^{-1} \) is a typical lifetime of the nucleus). If the conditional probability to decay is given by (12), then the conditional probability to survive during \( \Delta t \) after surviving during \( t \) is

\[
P_{\text{surv}}(\Delta t/t) = 1 - P_{\text{decay}}(\Delta t/t) = 1 - \gamma \Delta t \quad (\Delta t \ll \gamma^{-1}).
\]

Now we decompose the event of survival during time \( t \) into the product of survivals during \( m \) intervals \( \Delta t = t/m \) (with \( m \) large enough to guarantee \( \Delta t \ll \gamma^{-1} \)) and use the product of conditional probabilities to find

\[
P_{\text{surv}}(t) = P_{\text{surv}}(\Delta t/t_1) \cdot P_{\text{surv}}(\Delta t/t_2) \cdots \cdot P_{\text{surv}}(\Delta t/t_m) = (1 - \gamma \Delta t)^m = (1 - \gamma t/m)^m \quad (m \to \infty). \]

Noting that

\[
\lim_{m \to \infty} (1 - \gamma t/m)^m = \lim_{m \to \infty} e^{m \ln(1 - \gamma t/m)},
\]

and Taylor-expanding the logarithm \([\ln(1 + x) \to x, \text{ at } x \to 0]\) we finally obtain

\[
P_{\text{surv}}(t) = e^{-\gamma t}. \]
Random variables

In Statistical Physics, stochastic events are often associated with some random variables. For example, the energy or/and the number of particles of some open system; or velocity of a given particle, or just the time of the event like the decay of radioactive nucleus. If random variable is discrete, like numbers on a die, or zero and one respectively associated with heads and tails of a coin, then we may speak of the probability of finding a particular value of random variable. If the variable $x$ is continuous, the probability of finding some particular value of $x$ is zero, and we speak only of probability of finding the value of $x$ within some measurable set $Q$, that is the probability of the event $x \in Q$. The key concept for a continuous random variable $x$ is the probability density (also called probability distribution), $w(x)$, which allows one to introduce the probability for the event $x \in Q$ as

$$P_{x \in Q} = \int_{x \in Q} w(x) \, dx .$$

(17)

Clearly,

$$\int w(x) \, dx = 1 ,$$

(18)

because each time we measure $x$, we find some value. The notion of probability distribution is a bit more general than the notion of probability density, as it is also applicable to discrete variables. In the case of discrete random variable—like the number on a die—one understands by probability distribution the set of all the probabilities, $\{P_j\}$, to get the values $x = x_j$. The analog of (18) then is

$$\sum_j P_j = 1 .$$

(19)

If we allow the probability density to be expressed in terms of $\delta$-functions, then the notions of probability distribution and density become identical, and we can use Eqs. (17)-(18) in a discrete case as well. This is very convenient, and from now on we do not distinguish between continuous and discrete distributions, unless the opposite is specified. Also, we can understand $x$ not only as a number, but also as a $n$-dimensional vector $x \equiv (x_1, x_2, \ldots, x_n)$, or, equivalently, a point in $n$ dimensional sample space. The integrals (17)-(18) are then understood as $n$-dimensional integrals. For example, in the section “Number $\pi$ by Monte Carlo” we are dealing with two cases of two-dimensional sample spaces. The crucial feature of both cases is that—by symmetry—the probability density is independent of the random vector. These are so-called flat distributions. Clearly, classical probabilities are always associated with flat distributions of simple events. In the case of continuous flat distribution, $w$ is equal to the inverse volume (area in 2D, length in 1D) of the $n$-dimensional region $R$ of allowed values of the vector $x$.—This is seen from (18).

**Problem 9.** Distribution of the time moment of the decay: At $t = 0$ we start observing a radioactive particle with the decay rate $\gamma$. What is the probability that the decay of the particle happens during the infinitesimal time interval $[t, t + dt]$?

Given a random variable $x$ and its distribution $w(x)$, we can find its expectation value

$$\langle x \rangle = \int x \, w(x) \, dx .$$

(20)

In accordance with the definition of statistical probability, the expectation value (20) is nothing else than the arithmetic mean value of $x$ in an infinitely large number of trials. Indeed, arithmetic mean of $x$ is defined as

$$\bar{x} = \frac{x(1) + x(2) + \ldots + x(N)}{N} .$$

(21)
Here \( x(j) \) is the value of \( x \) in the \( j \)-th trial. We can rewrite this as (to simplify the notation we treat \( x \) as a discrete variable)

\[
\bar{x} = \frac{X_1N_1 + X_2N_2 + \ldots + X_mN_m}{N},
\]

(22)

where \( X_i \) is the \( i \)-th possible value of \( x \) and \( N_i \) is the number of events \( x = X_i \). For a very large \( N \) we thus have

\[
\bar{x} = \sum_{i=1}^{m} X_i(N_i/N) \rightarrow \sum_{i=1}^{m} X_iP_i \rightarrow \int x w(x) \, dx \equiv \langle x \rangle.
\]

(23)

Assuming \( N \) to be very large, from now on we set \( \bar{x} = \langle x \rangle \).

It is also important to know how large are the fluctuations of the values of \( x \) from \( \langle x \rangle \) in a large number of trials. These fluctuations can be characterized by standard deviation, \( \Delta x \), which is a non-negative quantity defined as

\[
(\Delta x)^2 = \int (x - \bar{x})^2 w(x) \, dx.
\]

(24)

We are often interested in expectations and standard deviations of not only \( x \) itself, but also of some function \( A(x) \). For example, if \( v \) is a random variable standing for a velocity of a particle interacting with some heat bath, then the random variable standing for energy is \( \propto v^2 \). Thus, we generalize Eqs. (20)-(24):

\[
\bar{A} = \langle A \rangle = \int A(x) w(x) \, dx,
\]

(25)

\[
(\Delta A)^2 = \int [A(x) - \bar{A}]^2 w(x) \, dx.
\]

(26)

In this sense,

\[
(\Delta x)^2 \equiv \langle (x - \bar{x})^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2,
\]

(27)

and, more generally,

\[
(\Delta A)^2 \equiv \langle (A - \bar{A})^2 \rangle = \langle A^2 \rangle - \langle A \rangle^2.
\]

(28)

[There is a synonym to the standard deviation: dispersion. Also, sometimes a shorter notation is used: \( \sigma \equiv \Delta x \), or \( \sigma_A \equiv \Delta A \).]

**Expectation of a sum of random variables.** It is easy to find an expectation value of a sum of two random variables (by just noting the linearity of the procedure):

\[
\langle x_1 + x_2 \rangle = \int (x_1 + x_2) W(x_1, x_2) \, dx_1 dx_2 = \int x_1 W(x_1, x_2) \, dx_1 dx_2 + \int x_2 W(x_1, x_2) \, dx_1 dx_2 = \bar{x}_1 + \bar{x}_2.
\]

(29)

By induction,

\[
\langle x_1 + x_2 + \ldots + x_m \rangle = \bar{x}_1 + \bar{x}_2 + \ldots + \bar{x}_m.
\]

(30)

**Independent random variables** play a very important role in Statistical Physics. A simple example is given by coordinates of non-interacting particles. More non-trivial example—which we will be dealing with later on in this course—is given by the momentum and coordinate of one and the same particle, provided the particle is in equilibrium with a heat bath.

Consider two independent random variables, \( x_1 \) and \( x_2 \), with their probability densities, \( w_1(x_1) \) and \( w_2(x_2) \). What is the distribution function \( W(x_1, x_2) \) for a pair \( (x_1, x_2) \)? The pair \( (x_1, x_2) \) can be visualized as a point in the plane \( x_1 x_2 \). By its definition, \( W(x_1^{(0)}, x_2^{(0)}) \) describes the probability \( dP = W(x_1^{(0)}, x_2^{(0)}) \, dx_1 dx_2 \) to find our random point in a rectangle \( dx_1 \times dx_2 \) at the point \( (x_1^{(0)}, x_2^{(0)}) \).
This event is a product of two events: (1) the event $x_1 \in [x_1^{(0)}, x_1^{(0)} + dx_1]$ and (2) the event $x_2 \in [x_2^{(0)}, x_2^{(0)} + dx_2]$. As the variables $x_1$ and $x_2$ are independent, the events (1) and (2) are independent as well, and we have

$$dP = dP_1 dP_2 ,$$

where $dP_1 = w_1(x_1^{(0)}) dx_1$ and $dP_2 = w_2(x_2^{(0)}) dx_2$ are the probabilities of the events (1) and (2).

For any pair $(x_1^{(0)}, x_2^{(0)})$ we thus have

$$W(x_1^{(0)}, x_2^{(0)}) dx_1 dx_2 = w_1(x_1^{(0)}) w_2(x_2^{(0)}) dx_1 dx_2 ,$$

that is

$$W(x_1, x_2) = w_1(x_1) w_2(x_2) .$$

By induction, we generalize Eq. (33) to any number $m$ of independent random variables:

$$W(x_1, x_2, \ldots, x_m) = w_1(x_1) \cdot w_2(x_2) \cdot \ldots \cdot w_m(x_m) .$$

With Eq. (33) we obtain a relation for the product of independent variables:

$$\langle x_1 x_2 \rangle = \int x_1 x_2 W(x_1, x_2) dx_1 dx_2 = \bar{x}_1 \bar{x}_2 .$$

And by induction we have,

$$\langle x_1 \cdot x_2 \cdot \ldots \cdot x_m \rangle = \bar{x}_1 \cdot \bar{x}_2 \cdot \ldots \cdot \bar{x}_m .$$

It is important to emphasize that Eqs. (35)-(36), in contrast to Eqs. (29)-(30), are valid only for independent variables.

Consider a random variable

$$y = x_1 + x_2 + \ldots + x_m ,$$

where all $x_j$’s are independent. Eq. (30) tells us that $\bar{y} = \bar{x}_1 + \bar{x}_2 + \ldots + \bar{x}_m$. But what about $\Delta y$? To find $\Delta y$ we once again use (33). First we consider $m = 2$, and find

$$(\Delta y)^2 = \iint (x_1 + x_2 - \bar{x}_1 - \bar{x}_2)^2 W(x_1, x_2) dx_1 dx_2 = \langle x_1^2 \rangle - \bar{x}_1^2 + \langle x_2^2 \rangle - \bar{x}_2^2 ,$$

which means that the square of dispersion of a sum is the sum of squares of dispersions:

$$(\Delta y)^2 = (\Delta x_1)^2 + (\Delta x_2)^2 .$$

By induction, this result generalizes to any $m$:

$$(\Delta y)^2 = (\Delta x_1)^2 + (\Delta x_2)^2 + \ldots + (\Delta x_m)^2 ,$$

This relation is very important. In particular, it tells us that if some random variable $y$ is a sum of a huge number of independent variables with finite dispersion, then its relative fluctuations are negligibly small. Indeed, $\bar{y}$ scales like $m$ as $m \to \infty$, while $\Delta y$ scales only like $\sqrt{m}$, hence

$$\frac{\Delta y}{\bar{y}} \propto \frac{1}{\sqrt{m}} \to 0 \quad \text{as} \quad m \to \infty .$$

Such a behavior is typical to additive macroscopic quantities like the total energy and the number of particles in a macroscopic open system.
Central Limiting Theorem

In the limit of \( m \gg 1 \), one can establish not just the relation (40), but the whole distribution function \( w(y) \) for the variable \( y \). The celebrated Central Limiting Theorem states that, in the limit of \( m \to \infty \), the distribution of the variable \( y \) is totally defined by its average value and dispersion, and is given by (below \( \sigma \equiv \Delta y \))

\[
w(y) = \frac{e^{-(y-\bar{y})^2/2\sigma^2}}{\sqrt{2\pi} \sigma}.
\] (42)

The distribution (42) is called Gaussian distribution. Note that the denominator of the Gaussian distribution is nothing else than a normalization factor following from the requirement \( \int w(y) \, dy = 1. \) It is unambiguously defined by the exponential. [The proof of the Central Limiting Theorem goes beyond the scope of this course.]

Binomial distribution

Suppose we perform \( N \) trials for some stochastic event \( A \) characterized by probability \( p \). What is the probability \( P(N_A, N) \) to have exactly \( N_A \) successful trials out of the total of \( N \)? The answer for \( P(N_A, N) \) is readily obtained (a slightly generalized problem 5, where instead of \( p = 1/2 \) we are dealing with an arbitrary \( p \)):

\[
P(N_A, N) = p^{N_A}(1-p)^{N-N_A} \frac{N!}{N_A!(N-N_A)!}.
\] (43)

Eq. (43) is called binomial distribution. Ultimately, we would like to understand the behavior of this distribution when \( N \gg 1 \). The expression Eq. (43) is not transparent enough to directly see this limit. (One has first to utilize the Stirling’s formula for the factorials, and then to work with the resulting expressions.) We can circumvent this difficulty by the following trick. Let us introduce a random variable \( x_j \) which takes on just two values: 1, if the event \( A \) happens in the \( j \)-th trial, and 0, if the event \( A \) does not happen in the \( j \)-th trial. We have

\[
\langle x_j \rangle = 0 \cdot (1-p) + 1 \cdot p = p.
\] (44)

\[
\langle x_j^2 \rangle = 0^2 \cdot (1-p) + 1^2 \cdot p = p.
\] (45)

\[
(\Delta x_j)^2 = \langle x_j^2 \rangle - \langle x_j \rangle^2 = p(1-p).
\] (46)

Then we notice that

\[
N_A = x_1 + x_2 + x_3 + \ldots + x_N,
\] (47)

where all \( x \)'s are independent variables, and utilize the results of the previous sections:

\[
\langle N_A \rangle = N \langle x_j \rangle = pN,
\] (48)

which is not at all a surprise, and

\[
\Delta N_A = \sqrt{N(\Delta x_j)^2} = \sqrt{Np(1-p)},
\] (49)

which is a really important result showing how the standard deviation behaves with \( N \).

At large \( N \) we can use the Central Limiting Theorem, Eq. (42), with \( y \equiv N_A, \bar{y} \equiv pN, \) and \( \sigma \equiv \Delta N_A = \sqrt{Np(1-p)}. \)

Poisson Distribution

Consider a detector (counter) of cosmic particles. Assume that the radiation background is constant.
in time and is completely characterized by the parameter $\xi$, the number of counts per unit time. Hence, during the time $t$ we expect $\xi t$ counts. The actual number of counts during the time $t$ can, however, differ from $\xi t$, due to fluctuations. Hence, this number is a random variable, and a natural question is: What is the probability $P_m(t)$ that this number equals $m$? Let us find $P_m(t)$. First, we slice up the time interval $t$ into $N > 1$ very small sub-intervals $\Delta t = t/N$, such that

$$P_0(\Delta t) \gg P_1(\Delta t) \gg P_2(\Delta t) \gg P_3(\Delta t) \gg \ldots .$$

(50)

In the limit $\Delta t \to 0$, we can neglect the cases when more than one counts take place during one time slice. This allows us to decompose the event of having exactly $m$ counts during time $t$ into simple events of having/not having counts during each given time slice. And this is exactly the situation we have discussed in the previous section—the binomial distribution of successful events. By the definition of the parameter $\xi$,

$$P_1(\Delta t) \to \Delta t \cdot \xi \quad (\Delta t \to 0).$$

(51)

We plug this into Eq. (43) and get

$$P_m(t) = (\Delta t \cdot \xi)^m (1 - \Delta t \cdot \xi)^{N-m} \frac{N!}{m!(N-m)!}.$$

(52)

We still need to do some work to get rid of $N$ and $\Delta t$:

$$\lim_{N \to \infty} (1 - \Delta t \cdot \xi)^{N-m} = \lim_{N \to \infty} e^{(N-m)\ln(1-\Delta t \cdot \xi)} = \lim_{N \to \infty} e^{-\xi t(N-m)/N} = e^{-\xi t},$$

(53)

$$\lim_{N \to \infty} \frac{(\Delta t \cdot \xi)^m N!}{m!(N-m)!} = (\xi t)^m \lim_{N \to \infty} \frac{N-m+1}{N} \cdot \frac{N-m+2}{N} \cdot \ldots \cdot \frac{N-1}{N} = (\xi t)^m.$$

(54)

Our final answer then is:

$$P_m(t) = e^{-\xi t} \frac{(\xi t)^m}{m!}.$$

(55)

Another physically important example of Poisson distribution deals with an element of volume, $V_0$, of a gas of non-interacting (=independent) particles. The question is: What is the probability to find exactly $m$ molecules of the gas inside the volume $V_0$? The answer reads:

$$P_m(V_0) = e^{-nV_0} \frac{(nV_0)^m}{m!},$$

(56)

where $n$ is the number density.—Do you see why?

The general form of the Poisson distribution is

$$P_m(a) = e^{-a} \frac{a^m}{m!} \quad (m = 0, 1, 2, 3 \ldots),$$

(57)

where $a$ is some dimensionless parameter. Eq. (57) is easily memorizable: Up to a natural normalization factor of $e^{-a}$, the value of $P_m(a)$ is nothing else than the $m$-th term of the Taylor’s expansion of the exponential $e^a$.

**Problem 10.** Make sure that the following relations hold true for the distribution (57):

$$\langle m \rangle = a,$$

(58)

$$\Delta m = \sqrt{a}.$$

(59)

Apply the Central Limiting Theorem to Eq. (57) at $a \gg 1$.

**Problem 11.** On the basis of the physical meaning of the Poisson distribution, argue that

$$P_0(a_1)P_m(a_2) + P_1(a_1)P_{m-1}(a_2) + P_2(a_1)P_{m-2}(a_2) + \ldots + P_m(a_1)P_0(a_2) = P_m(a_1 + a_2).$$

(60)

[In this problem, I am not interested in the formal derivation of (60) from (57).]