## Contents

1 Random variables ................................................................. 2
   1.1 Some intuitive concepts .................................................. 2
   1.2 Probability and frequency ................................................. 3
      1.2.1 Properties of the PDFs .............................................. 4
   1.3 Probability of independent events ....................................... 5
      1.3.1 Problem: the birthday “paradox” .................................. 6
   1.4 Joint, disjoint, conditional probability .................................. 6
   1.5 Probability distributions .................................................. 8
   1.6 Variable transformations .................................................. 10
   1.7 Error propagation .......................................................... 11
      1.7.1 Sum and products of variables ..................................... 11
   1.8 The main PDFs .............................................................. 12
      1.8.1 Binomial PDF ......................................................... 12
      1.8.2 Poissonian PDF ..................................................... 13
      1.8.3 Gaussian PDF ....................................................... 14
Chapter 1

Random variables

1.1 Some intuitive concepts

Let us consider a group of $N = 120$ people and let’s record how many of these people were born in a given month. Suppose we obtain:

<table>
<thead>
<tr>
<th></th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_i$</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>11</td>
<td>7</td>
<td>13</td>
<td>5</td>
<td>9</td>
<td>7</td>
<td>10</td>
<td>8</td>
<td>15</td>
</tr>
</tbody>
</table>

The number of people in every $i$-th month, $n_i$, is a random variable, i.e. a variable that can assume random (unpredictable) values in a certain range. The theory of probability describes quantitatively the behavior of random variables.

The first intuitive concept to quantify the behavior of $n_i$ is to define an average:

$$\hat{n} \equiv \frac{\sum_n n_i}{s}$$

where $s = 12$ is the number of months in the year. If we don’t have any reason to suspect that some month is somehow special (we neglect here the slight difference in month lengths), we expect that every month contains $N/s$ people and since $N = \sum n_i$ we can define the average as above. The average itself may or may not be a random variable. In the present case, $\hat{n} = 10$ is not a random variable, since we fixed the total number of persons in our experiment. In most cases, e.g. in physical measurements, the average of a sample of measures is itself a random variable.

We ask now ourselves how to define the deviation from the mean, i.e. how to quantify the fact that generally $n_i \neq \hat{n}$. Perhaps we are interested in discovering anomalies in the distribution. If we are now not interested in the sign of the deviation but only on its amplitude we could use something proportional to $|n_i - \hat{n}|$. We could define then something like

$$\sum \frac{|n_i - \hat{n}|}{s}$$

However, for reasons that will be clear later one, usually we define a related but different quantity:

$$\sigma \equiv \sqrt{\frac{\sum_n (n_i - \hat{n})^2}{s}}$$

where $\sigma$ is called the root mean square or standard deviation (or more exactly an estimate of the standard deviation). Note that $\sigma^2$ is itself an average:

$$\sigma^2 \equiv \frac{\sum_n (n_i - \hat{n})^2}{s}$$

Sometimes we also use the notation

$$\hat{x} \equiv <x> \equiv \frac{\sum n_i x_i}{s}$$

In our present example we have then

$$\hat{n} = 10 \quad \sigma = 3.41$$

We expect then that most data $n_i$ will not deviate more than a few $\sigma$ from average dal and this is in fact what we observe. Then $\hat{n}$ and $\sigma$ give us important information on the behavior of $n_i$; since this experiment can be repeated many times, these averages describe fundamental properties of the data distribution. It is intuitively clear that these averages will be more and more precise when we average over more and more experiments. These ideas will be more precisely defined in the following.
Every function of $n_i$ is itself a random variable (although it could be a trivial random variable with probability 1). For instance, the number of people born in the first $M$ months; the number of months with more than $P$ people and so on.

### 1.2 Probability and frequency

Let us build now a data histogram, i.e., bars with height equal to the number $p_i$ of months with $n_j$ people, as in Fig. 1.1.

The bar height gives an estimate of the probability of obtaining that number $n_i$ of people. If we repeat the experiment $N_{exp}$ times the estimate will be more precise, in the sense that it will converge to a limiting case. For instance, in Figs 1.2, 1.3 we have the cases $N_{exp} = 10$ and 50.
For $N_{\text{exp}} \to \infty$ we can assume we have the “true probability” of having $n_i$. We define then

$$P(n_i) \equiv \lim_{N_{\text{exp}} \to \infty} \frac{\text{number of occurrences of } n}{sN_{\text{exp}}} \quad (1.7)$$

that is, the frequency of events. Now since $\sum \text{number of occurrences of } n = sN_{\text{exp}}$, we have

$$\sum_{i} P(n_i) = 1 \quad (1.8)$$

The sum of all possible probabilities of any given experiments equals 1. La Fig. 1.3 diventa allora come in Fig. 1.4. The histogram approximates the probability distribution of $n_i$.

In the limit in which the random variable $n_i$ becomes a continuous variable (eg a temperature, a magnitude etc), we define a probability density or probability distribution function (PDF) $f(x)$

$$f(x)dx = P(x) \quad (1.9)$$

and we have, within the domain of $x$ (i.e. all its possible values)

$$\int f(x)dx = 1 \quad (1.10)$$

In the same experiment we can identify other random variables and therefore other PDFs. For instance, if we ask the birth month we will get an answer in the range $m_i \in 1 − 12$ which itself is a random variable. Then we can create an histogram as in Fig. 1.5 in which we plot the month frequency $n_i/N_{\text{tot}}$ versus the months. If we increase $N_{\text{tot}}$ to 1200 we obtain Fig. 1.6. Here clearly the distribution tends to a uniform distribution with $P(m_i) = 1/12$.

### 1.2.1 Properties of the PDFs

The two most fundamental properties of probability distributions are

$$\int f(x)dx = 1 \quad (1.11)$$
We can easily extend the idea to joint events, for instance the probability of obtaining at the same time (non necessarily in the chronological sense) the measurement \( x \) in \( dx \) (eg a galaxy magnitude) and \( y \) in \( dy \) (eg the galaxy redshift). Then we have

\[
f(x, y)dx dy = P(x, y)
\]

\[
f(x, y) \geq 0
\]

\[
\int f(x, y)dx dy = 1
\]

Immediate consequence of the first law is that if \( F(<X) = \int_{-\infty}^{X} f(x)dx \) is the probability of obtaining a result less than \( X \), then the probability of obtaining a result greater than or equal to \( X \) is \( F(\geq X) = 1 - F(<X) \). So in general if \( P(A) \) is the probability of \( A \), the probability of non-\( A \) (ie anything but \( A \)) is simply \( 1 - P(A) \), to be denoted as \( P(\bar{A}) \).

Other examples of probability:

\[
P(x) = \lim_{N \to \infty} \frac{\text{number of people voting for party } X}{\text{number of interviewed}}
\]

\[
P(x) = \lim_{N \to \infty} \frac{\text{number of measures (distances, temp., etc) that give } X}{\text{number of experiments}}
\]

Clearly if \( x \) is a continuous variable we have

\[
f(x)dx = \lim_{N \to \infty} \frac{\text{number of measures in } x,x+dx}{\text{number of experiments}}
\]

### 1.3 Probability of independent events

Suppose we throw two dice; the joint probability of obtaining 1 in a throw and 2 in the other one is the product of the single-throw probabilities, \( P_{12} = P_1P_2 \). This is true only because the two throws are independent (do not influence each other). Then we have \( P_1 = P_2 = 1/6 \) and \( P_{12} = 1/36 \), as of course one could see by the number of occurrences over the number of experiments. If we have the PDF \( f(x, y) \), the event \( x \) can be said to be independent of event \( y \) if the probability of \( x \), \( p(x)dx \), does not depend on \( y \). Now the probability of having \( x \) in \( dx \) when \( y \) is in a sub-range \( \Delta_y \) is

\[
p(x)dx = \int_{\Delta_y} dy f(x,y)dx
\]

So in order for \( p(x) \) not to depend on \( y \) it is necessary that \( f \) be separable:

\[
f(x,y) = q(x)g(y)
\]

So in this case

\[
p(x)dx = \int_{\Delta_y} g(y)dy \int_{\Delta_x} q(x)dx = Nq(x)dx
\]
In general therefore not probability that the 3 birthdays does have: probability of non-coinc. is \( e^{-N} \) and therefore \( N \). We can now use \( N \). It is clear then that for mutually exclusive, we can write \[ P(x, y) = 0 \]

Finally, the probability of having at least one coincidence must be the complement to unity to this, i.e.

\[ P(non - coinc, N) = e^{-1/365}e^{-2/365}e^{-3/365}...e^{-(N-1)/365} = e^{-\frac{(N-1)}{365}} \]

(1.25)

Finally, the probability of having at least one coincidence must be the complement to unity to this, i.e.

\[ P(coinc, N) = 1 - e^{-\frac{N(N-1)}{365}} \approx 1 - e^{-\frac{2}{365}} \]

(1.26)

For \( N = 20 \) one has, perhaps surprisingly (this is the “paradox”) \( P(N) = 0.5 \) i.e. almost 50%.

1.3.1 Problem: the birthday “paradox”

Let us estimate the probability that in \( N \) random people there are at least two with the same birthday.

A person \( B \) has the same birthday of person \( A \) only once in 365. Then \( P(coinc., N = 2) = 1/365 \) and the probability of non-coinc. is \( e^{P(non - coinc., N = 2)} = 1 - 1/365 = 364/365 \).

Let’s add a third person. His/her birthday will not coincide with the other two 363 times over 365. The joint probability that the 3 birthdays does not coincide is then

\[ P(non - coinc., N = 3) = \frac{364}{365} \frac{363}{365} \frac{362}{365} \]

It is clear then that for \( N \) persons we have

\[ P(non - coinc., N) = \frac{365}{365} \frac{364}{365} \frac{363}{365} \ldots \frac{365 - N + 1}{365} \]

(1.23)

We can now use

\[ e^{-x} \approx 1 - x \]

(1.24)

to write

\[ \frac{365 - N + 1}{365} = 1 - \frac{N - 1}{365} \approx e^{-(N-1)/365} \]

and therefore

\[ P(non - coinc, N) = e^{-1/365}e^{-2/365}e^{-3/365}...e^{-(N-1)/365} = e^{-\frac{(N-1)}{365}} \]

(1.25)

1.4 Joint, disjoint, conditional probability

We can define three kind of probabilities.

Joint \( P \). If \( P_A \) and \( P_B \) are the probabilities of the independent events \( A \) and \( B \), the probability of having both \( A \) and \( B \) is \( P_A P_B \). Then

\[ P(A \cap B) = P(B \cap A) = P(A)P(B) \]

(1.27)

For instance, the prob. of having 1 in a dice throw and 2 in another one is \((1/6)^2 = 1/36\).

Disjoint \( P \). If \( P_A \) and \( P_B \) are the prob. of events \( A \) and \( B \) mutually exclusive (i.e. \( A \cap B = A \cap B = 0 \)), the prob. of \( A \) or \( B \) is \( P_A + P_B \). Therefore

\[ P(A \cup B) = P(A) + P(B) \]

(1.28)

We have already seen an example of disjoint prob. when we have seen that \( P(A) = 1 - P(\overline{A}) \). Since \( A \) and \( \overline{A} \) are mutually exclusive, we can write

\[ P(A \cup \overline{A}) = 1 = P(A) + P(\overline{A}) \]

(1.29)

So for instance the prob. of having 1 or 2 in a dice roll is \( 1/6 + 1/6 = 1/3 \). Considering continuous variables we have

\[ p(x \in A \cup B) = \int_A f(x)dx + \int_B f(x)dx \]

(1.30)

only if the ranges \( A \) and \( B \) do not overlap. If they overlap, the events are not mutually exclusive \( A \cap B \neq 0 \), and we have:

\[ p(x \in A \cup B) = \int_A f(x)dx + \int_B f(x)dx - \int_{A \cap B} f(x)dx \]

(1.31)

In general therefore

\[ P(A \cup B) = P(A) + P(B) - P(A \cap B) \]

(1.32)
So for instance if \( A \) is the prob. of having one “1” in the first die, whatever the second is, and \( B \) the prob. of “1” in the second die, whatever the first is, and we consider the prob. of having at least a “1” in two throws, the event “1” is both \( A \) and \( B \). So we have \( P(A \cup B) = 1/6 + 1/6 - 1/36 = 11/36 \), as we can verify easily since the winning combinations are \((11,12,13,14,15,16,21,31,41,51,61)\) are 11 over 36.

*Conditional P.* If the events are not independent, we can define the conditional probability (prob. of \( A \) given \( B \)):

\[
P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\text{number of cases that are both } A \text{ and } B}{\text{number of cases that are } B} \tag{1.33}
\]

So for instance, the probability of the combination 1-2 after obtaining 1 in the first roll equals \((1/36)/(1/6) = 1/6\).

This extends obviously to continuous variables. The prob. of \( x \) in the range \( I = (-1,1) \) given that \( x < 0 \) is \( P(x \in I|x < 0) \). The prob. of having \( x < 0 \) is

\[
P(x < 0) = \int_{-\infty}^{0} f(x) dx \tag{1.34}
\]

and the prob. of having \( x \in I \) and at the same time \( x < 0 \) is

\[
P(x < 0, x \in I) = \int_{-1}^{0} f(x) dx
\]

Now, the fraction of cases (or area) such that \( P(x \in I|x < 0) \) is clearly the fraction \( P(x < 0, x \in I)/P(x < 0) \), which agrees with the rule above. In other words, if in 100 measures there are 50 with \( x < 0 \) and 20 with \(-1 < x < 0 \) it is clear that the fraction of measures with \( x \in I \) among those with \( x < 0 \) is \( 20/50 = 2/5 \).

Another example. The prob. of obtaining \( \geq 9 \) in two dice rolls is \( 10/36 \): there are in fact 10 successful events: 36, 45, 46, 55, 56, 66, 63, 54, 64, 65 in 36 possibilities. Which is the prob. of obtaining a score \( \geq 9 \) given that in the first roll the result is 6 ? We have

\[
P(x + y \geq 9|x = 6) = P(x + y \geq 9, x = 6)/P(x = 6) = \frac{4}{36} = \frac{2}{3} \tag{1.35}
\]

which indeed is true since if the first die has a 6, then it is sufficient that the second result is 3,4,5,6 to win, i.e. 4 cases out of 6.

Yet another example. Suppose we know that on average one person out of 1000 randomly chosen ones is a physics student and plays piano; then we also know that in general one out of 100 people plays piano; then the probability that a person, among those that play piano, is also a physics student is \((1/1000)/(1/100) = 1/10\). In other words, out of 1000 people, we know that 10 (on average) play piano; we also know that among those 1000 people there is one that both plays piano and study physics and of course this person has to be among the 10 that play piano. Then this person is indeed 10% of those that play piano.

Consequently, the fraction of people that play piano and study physics, \( P(A \cap B) \), is equal to the fraction that play piano, \( P(B) \), times the fraction of people that study physics among those that play piano, \( P(A|B) \). Now, since, \( A \cap B = B \cap A \) we have Bayes’ Theorem

\[
P(A \cap B) = P(A|B)P(B) = P(B|A)P(A) \tag{1.36}
\]

or

\[
P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{1.37}
\]

Note that if \( A \) and \( B \) are independent, ie if

\[
P(A \cap B) = P(A)P(B) \tag{1.38}
\]

it follows that \( P(A|B) = P(A) \) and \( P(B|A) = P(B) \). For instance, if the fraction of people that study physics is \( 1/100 \) and the fraction that play piano is \( 1/100 \), if we find out that the fraction of people that both study physics and play piano is not \( 1/10000 \) then we can conclude that playing piano and studying physics are not independent events. The prob. \( P(A) \) in this case is called *prior probability*; the prob. \( P(A|B) \) is called *posterior probability*.

The prob. that \( B \cup \overline{B} \) occurs is of course always 1, even in the case of conditional prob. We have therefore

\[
P(B \cup \overline{B}) = 1 = \frac{P(A,B \cup \overline{B})}{P(A)} \tag{1.39}
\]

or

\[
P(A,B \cup \overline{B}) = P(A) \tag{1.40}
\]
In terms of PDF this rule says that integrating a PDF of two variables over the whole domain of one of the two (marginalization) we obtain the PDF of the other:

$$\int f(x,y)dy = p(x) \quad (1.41)$$

Clearly if $f(x,y) = p(x)g(y)$, the $p(x)$ is indeed the PDF of $x$, which of course confirms that a factorizable $f(x,y)$ implies that $x$ and $y$ are independent.

**Problem.**

1% of people has the tropical disease $Z$. There exist a test that gives positive 80% of the times if the disease is present (true positive), but also 10% of the times when the disease is absent (false positive). If a person tests positive, which is the prob. that he/she has the $Z$ disease?

**Answer.** We have:

- prob. of having $Z$: $P(Z) = 1\%$
- cond. prob. of being positive (event labeled $p$) having $Z$: $P(p|Z) = 80\%$
- cond. prob. of being positive while not having $Z$: $P(p|\overline{Z}) = 10\%$.

From the first we deduce that $P(Z) = 99\%$. We need now the prob. of having $Z$ being positive; from Bayes’ theorem we have:

$$P(Z|p) = \frac{P(p|Z)P(Z)}{P(p)} \quad (1.42)$$

We should then evaluate $P(p)$. The prob. of testing positive and having $Z$ is:

$$P(p, Z) = P(p|Z)P(Z) = 0.8 \cdot 0.01 = 0.008 \quad (1.43)$$

The prob. of testing positive being healthy is instead:

$$P(p, \overline{Z}) = P(p|\overline{Z})P(\overline{Z}) = 0.1 \cdot 0.99 \approx 0.1 \quad (1.44)$$

Moreover

$$P(p) = P(p, Z) + P(p, \overline{Z}) = P(p|Z)P(Z) + P(p|\overline{Z})P(\overline{Z}) = 0.108$$

It follows finally

$$P(Z|p) = \frac{P(p|Z)P(Z)}{P(p)} = \frac{0.8 \cdot 0.01}{0.108} = 0.075 \quad (1.45)$$

ie there is only a prob. of 7.5% of having $Z$. The reason of this perhaps surprising result is that $P(Z)$ (the absolute prob. of being infected with $Z$) is much smaller than $P(p)$, the absolute prob. of testing positive.

Exchanging the conditional probabilities is a very common logical error: the probability that one is a great artist because he/she is "misunderstood" (ie, nobody likes his/her paintings) is not equal to the probability of being misunderstood being a great artist. In fact, the probability of being great artists is much less than the probability of making bad paintings.

**Exercise:** what is the probability of throwing two dice and obtain a score between 1 and 3 in the first roll and 6 in the second? what is the probability of having an overall score less than 8 in the launch of two dice? And if we had already obtained 5 in the first die?

### 1.5 Probability distributions

Let’s briefly introduce two examples of PDFs.

Uniform distribution.

$f(x) = const.$ in the range $x \in (a - b)$. We have

$$\int_a^b f(x)dx = const. \times (b - a) \quad (1.46)$$

and the normalization requires $\text{const} = 1/(b - a)$.

Gauss distribution.

$$f(x) = A e^{-\frac{(x-x_0)^2}{2\sigma^2}} \quad (1.47)$$
Normalization
\[ \int f(x)\,dx = A \int_{-\infty}^{+\infty} \exp(-\frac{(x-x_0)^2}{2\sigma^2})\,dx = A\sqrt{2\pi\sigma^2} \] (1.48)
from which \( A = (2\pi\sigma^2)^{-1/2} \).

Indeed:
\[
\int e^{-x^2/2}\,dx = \sqrt{\int e^{-x^2/2}\,dx} \int e^{-y^2/2}\,dy \\
= \sqrt{\int e^{-(x^2+y^2)/2}\,dxdy} \\
= \sqrt{\int e^{-r^2/2}rdr \int_{-\pi}^{+\pi} d\theta} \\
= \sqrt{2\pi \int_0^{+\infty} e^{-z}\,dz} = \sqrt{2\pi(-e^{-\infty} + e^0)} = \sqrt{2\pi}
\]

Finally since
\[
\int e^{-x^2/2\sigma^2}\,dx = \sigma \int e^{-z^2/2}\,dz = \sigma \sqrt{2\pi}
\]
we obtain the result. The parameters \( x_0 \) and \( \sigma^2 \) are called mean and variance.

PDFs are characterized also by other quantities.

**Quantile \( \alpha \):** value of \( x \) such that
\[
\int_{-\infty}^{x} f(x')\,dx' = \alpha \quad (0 \leq \alpha \leq 1).
\]
If \( \alpha = 0.5 \) the quantile is called median.

**Mode.** The value of \( dx \) such that \( P(x) \) is maximal.

**Moments or expected values.**

The expected value of a quantity \( g(x) \) is defined as
\[
E[g] = \langle g \rangle = \hat{g} = \int g(x)f(x)\,dx
\] (1.50)
The mean is therefore the expectation value of \( x \):
\[
E[x] = \int x f(x)\,dx
\] (1.51)
For discrete variables we have
\[
E[n] = \sum_{i=1}^{N} n_i P(n_i)
\] (1.52)
Since \( P(n_i) \) is defined as the number of events \( n_i \) divided by the total number of cases, we retrieve the intuitive definition of mean of a variable as the sum of all the values divided by the number of cases.

The variance (or central moment of second order) is defined as
\[
E[(x - \hat{x})^2] = \int (x - \hat{x})^2 f(x)\,dx = \int x^2 f(x)\,dx - \hat{x}^2
\] (1.53)
For a Gaussian one has
\[
E[x] = x_0 \quad (1.54) \\
E[(x - \hat{x})^2] = \sigma^2 \quad (1.55)
\]
Note that \( E[x - \hat{x}] = 0 \) and \( E[y^2] \neq E[y]^2 \).

The variance has great importance in scientific measures. Conventionally in fact the error associated to each measure is given by the square root of the variance, or standard deviation, and is denoted generally with \( \sigma \) also for
non-Gaussian distributions. Note that $E^{1/2}[(x - \bar{x})^2]$ coincides indeed with the definition (eq. 1.4) in the limit of an infinite number of observations.

The $n$-th order moment is

$$E[x^n] = \int x^nf(x)dx$$ (1.56)

$$E[(x - \bar{x})^n] = \int(x - \bar{x})^nf(x)dx$$ (1.57)

Exercises:

Evaluate $E[x]$ and $E[(x - \bar{x})^2]$ for a uniform distribution in the range $(a - b)$ and for a Gaussian. Invent a PDF, normalize it, and evaluate mean and variance.

Prove that

$$E[ax] = aE[x]$$ (1.58)

$$E[x + a] = E[x] + a$$ (1.59)

that is, the mean is a linear operation.

### 1.6 Variable transformations

Given a random variable $x$ and its PDF $f(x)$, we could be interested to derive a PDF of a variable function of $x$, for instance $x^2$ or $1/x$ or $y(x)$. For instance, if we know the PDF of the absolute magnitude $M$ of a galaxy, we could be interested in the PDF of its distance $r$

$$M = m - 25 - 5\log_{10}r$$ (1.60)

assuming we know the apparent magnitude $m$. If $dy = y'dy$ is the infinitesimal interval of the new variable $y$ as a function of the old one, it is clear that the prob. of having $x$ in $x, x + dx$ must be equal to the one of having $y$ in $y, y + dy$:

$$f(x)dx = g(y)dy$$ (1.61)

and therefore the new PDF $g(y)$ is

$$g(y) = f(x)|dy/dy|$$ (1.62)

where the absolute value ensures the positivity of the new PDF. So if the PDF of $M$ is a Gaussian, the PDF of $r(M)$ is

$$f(r) = A|dr/dM|\exp - \frac{(M(r) - M_0)^2}{2\sigma^2} = A(\exp 10)^{5\log_{10}r} \exp - \frac{(m - 25 - 5\log r - M_0)^2}{2\sigma^2}$$

and defining $r_0$ such that $M_0 = m - 25 - 5\log r_0$ one can write

$$f(r) = A'\exp - \frac{(5\log r/r_0)^2}{2\sigma^2}$$ (1.63)

called a log-normal distribution. Notice that in general

$$E[g(x)] \neq g(E[x])$$ (1.64)

We can also consider the transformation of variables in the case of many random variables. The transformation from $x_1, x_2, \ldots$ to $y_1, y_2, \ldots$ can be performed introducing the Jacobian of the transformation

$$f(x_i)d^n = g(y_i)dy$$ (1.65)

from which

$$g(y_i) = f(x_i)|J|$$ (1.66)

where $J_{ij} = \partial x_i/\partial y_j$ and $||$ denotes the determinant.

**Exercise.**

If the variable $x$ is distribuita in a uniform manner in $(a, b)$, which is the distribution of $y = x^2$?
1.7 Error propagation

We can now use these formulae to find the error (standard deviation) associated to a function of a random variable \( x \) in the limit of small deviations from the mean.

Suppose we have a variable \( x \), eg the side of a square, distributed as \( f(x) \) with mean \( \mu \) and variance \( \sigma_x^2 \) and we are interested in the PDF of the area \( y = x^2 \). We can expand \( y \) around \( \mu \):

\[
y(x) = y(\mu) + \frac{dy}{dx}|_{\mu}(x - \mu)
\]

We can then evaluate the mean and variance in the limit of small deviations from \( \mu \):

\[
E[y] = \int y(\mu)P(x)dx = y(\mu) \int f(x)dx = y(\mu)
\]

and

\[
E[y^2] = \int [y^2(\mu) + y'(\mu)^2(x - \mu)^2 + 2y(\mu)y'(\mu)(x - \mu)]f(x)dx
\]

\[
= y^2(\mu) + y^2\sigma_x^2
\]

(where \( y' \equiv \frac{dy}{dx}|_{\mu} \)). It follows that the variance of \( y \) for small deviations of \( x \) from \( \mu \) is

\[
\sigma_y^2 = E[(y - y(\mu))^2] = E[y^2] - y(\mu)^2 = y^2\sigma_x^2
\]

In the case of area \( y = x^2 \) we have then \( \sigma_y^2 = 4\mu^2\sigma_x^2 \). This is the fundamental rule of error propagation.

We can easily extend this rule to several variables. Suppose for instance that \( y(x_1, x_2) \) depends on two variables, for instance \( y \) is the sum of the sides of two squares measured independently. Because of independence \( f(x_1, x_2) = f_1(x_1)f_2(x_2) \). Then we have

\[
y(x_1, x_2) = y(\mu_1, \mu_2) + \sum_i \frac{\partial y}{\partial x_i}|_{\mu}(x_i - \mu_i)
\]

from which

\[
E[y^2] = \int [y^2(\mu_1, \mu_2) + \sum_i y'_i(\mu)^2(x_i - \mu)^2 + 2y(\mu_1, \mu_2) \sum_i y'_i(\mu)(x_i - \mu)]f_1(x_1)f_2(x_2)dx_1dx_2
\]

\[
= y^2(\mu_1, \mu_2) + \sum_i y^2\sigma_{x_i}^2
\]

and finally

\[
\sigma_y^2 = E[(y - y(\mu_1, \mu_2))^2] = \sum_i y_i^2\sigma_{x_i}^2
\]

This rule extends obviously to any number of independent variables.

1.7.1 Sum and products of variables

In the case \( y = x_1 + x_2 + \ldots + x_n \) the above rule gives

\[
\sigma_y^2 = \sum_i \sigma_{x_i}^2
\]

i.e., the variance of a sum of random variables is the sum of the variances. The error in \( y \) is therefore

\[
\sigma_y = \sqrt{\sum_i \sigma_{x_i}^2}
\]

i.e. the errors add in quadrature.

Esercizio: generalize to \( y = a_1x_1 + a_2x_2 + \ldots + a_nx_n \).

In the case of a product, \( y = x_1x_2\ldots x_n \) we have instead

\[
\frac{\sigma_y^2}{y^2} = \sum_i \frac{\sigma_{x_i}^2}{x_i^2}
\]
where \( y = \mu_1 \mu_2 \ldots \mu_n \) e \( \hat{x}_i \equiv \mu_i \). The quantity
\[
\sigma_x
\]
is the relative error. For a product of variables, then, the square of the relative error is the sum of the squares of the individual relative errors. That is, for a product of variables the relative errors add in quadrature.

**Esercizio:** generalize to \( y = x_1^{n_1} x_2^{n_2} \ldots x_m^{n_m} \).

### 1.8 The main PDFs

#### 1.8.1 Binomial PDF

Let us consider \( N \) events, eg the scores 1–3–2–6 etc in a series of dice rolls, or the sequence TTCCCT of heads/crosses in coin tosses. We want to evaluate the probability that a joint event, eg 8 heads out of 10 tosses, or three times a 1 out of 4 dice rolls, regardless of the order in the sequence, i.e. considering the events as indistinguishables. This is exactly the same kind of statistics we need in eg the statistics of a gas, which depends on the probability for indistinguishable particles to be in a given region of phase space.

We need first of all to evaluate the number of possible sequences. If we have \( N \) different elements, ag \( a, b, c, \) we can permutate the \( N \) elements \( N! \) times. For instance, \( N = 3 \) elements can be combined \( 3! = 6 \) times: \( abc, acb, cab, cba, bac, bca. \) Then\( N! \) is the number of permutations of distinguishable elements.

Suppose now we have only two elements, eg head or cross, or event \( A \) and any other event \( \bar{A} \). Then many permutations are identical, for instance \( HHCC \) remains the same by exchanging the two \( H \)s and the three \( C \)s. Suppose we have \( N \) times one of the two elements and, therefore, \( N - n \) the number of the other. Then, among the total \( N! \) permutations, a fraction \( n! \) is identical because we permute the same identical \( n \) element, and a fraction \( (N - n)! \) will also be identical for the same reason. How many indistinguishable combinations will we obtain? Clearly
\[
\frac{\text{total permutations}}{\text{(permutations among \( n \))}(\text{permutations among \( N - n \))}} = \frac{N!}{n!(N-n)!} \equiv \binom{N}{n}
\]

For instance, if \( N = 4 \) and \( n = 2 \) (as in TTCC) I will have \( 4!/2!/2! = 6 \) equivalent combinations (\( HHCC, HCHC, CCHH, CHCH, CHHC, HCCH \)). Notice that for \( n = 0 \) we define \( n! = 1 \).

The binomial PDF generalizes this calculation to the case in which I have a series of \( n \) independent events \( A \) each with the same probability \( p \) (eg for “head” the prob. is \( 1/2 \), for a 2 in a dice roll is \( 1/6 \) etc). In this case, the occurrence of \( n \) events \( A \) or prob. \( p \) out of \( N \) implies the occurrence of \( N - n \) events \( \bar{A} \) with prob. \( 1 - p \). All this implies a joint prob. of
\[
p^n (1-p)^{N-n}
\]
But clearly we have \( \binom{N}{n} \) of such combinations and therefore the binomial prob. will be
\[
P(n; N, p) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}
\]
where \( n \) is the discrete random variable \( 0 \leq n \leq N \) (number of events \( A \)) while \( N, p \) are the distribution parameters.

Notice that by employing the rules of the binomial we have, as indeed we should have expected:
\[
\sum_n P(n; N, p) = (p + (1-p))^N = 1
\]
It is also intuitive that the mean of events \( A \) of prob. (frequency) \( p \) out of \( N \) events should be the fraction \( p \) of \( N \) and indeed
\[
E[n] = Np
\]
\[
\sigma^2 = E[(n - Np)^2] = Np(1-p)
\]
Let’s demonstrate the first one:
\[
E[n] = \sum_n n P(n; N, p) = \sum_{n=0}^{N} \frac{nN!}{n!(N-n)!} p^n (1-p)^{N-n}
\]
\[
= \sum_{n=1}^{N} \frac{N(N-1)!}{(n-1)!(N-n)!} pp^{n-1}(1-p)^{N-n}
\]
\[
= Np \sum_{n'=0}^{N'} \frac{(N')!}{(n')!(N' - n')!} p^{n'} (1-p)^{N'-n'} = Np
\]
Figure 1.7: Binomial for $N = 120$ and $p = 1/12$ (red dots) e $p = 1/24$ (blue dots).

Exercises:
1) Which is the probability of obtaining two heads out of 4 throws?
The prob. of having exactly $n = 2$ two heads, each with prob. $p = 0.5$, out of $N = 4$ events is

$$P(2; 4, 0.5) = 3/8$$ (1.87)

2) In the birthday experiment we have obtained 15 persons in December. Which is the prob. of obtaining 15 or more in a given month?
The prob. that the birthday of a person is in December is $p = 1/12$. The total number of events is $N = 120$ and $n = 15$ is the number of events $A =$ December. The statistics is then a Binomial $P(15; 120, 1/12)$. The prob. of having more than 15 events $A$ is therefore

$$F(>15) = \sum_{n>15} \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n} = 0.074$$

(1.88)

ovvero solo il 7.4%.

1.8.2 Poissonian PDF
Let us consider now the limit of the Binomial for $N \to \infty$ and $p \to 0$ (rare events), but keeping $Np = \nu$ finite. We can approximate $N!/(N-n)! \approx N^n$ and $(1-p)^{N-n} \approx e^{-n(1-p)} \approx e^{-\nu}$

$$P(n; \nu) = \frac{N^n}{n!} p^n e^{-\nu} = \frac{\nu^n}{n!} e^{-\nu}$$

(1.89)

and we obtain the Poissonina PDF.
The moments are

$$E[n] = e^{-\nu} \sum n \frac{\nu^n}{n!} = \nu$$

(1.90)

$$E[(n-\nu)^2] = \nu$$

(1.91)

For large $n$, we can assume that $n$ is a continuous variable. In this case we generalize to

$$P(x; \nu) = \frac{\nu^x}{\Gamma(x+1)} e^{-\nu}$$

(1.92)

where $\Gamma(x)$ (equal to $(x-1)!$ for $x$ integer) is the gamma function

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt$$

(1.93)
1.8.3 Gaussian PDF

For large $\nu$, the Poissonian is well approximated by the Gaussina with mean and variance $\nu$. The Gaussian is defined as:

$$G(\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$  \hfill (1.94)

and has mean $\mu$ and variance $\sigma^2$. Defining the new variable $z = (x - \mu)/\sigma$ the Gaussina becomes the Normal distribution:

$$N(x) = G(0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$  \hfill (1.95)

We can define the error function

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$  \hfill (1.96)

so that the cumulative function $F(x) = \int_{-\infty}^x G(x; 0, 1) dx$ becomes

$$F(x) = \frac{1}{2} [1 + erf\left(\frac{x}{\sqrt{2}}\right)]$$  \hfill (1.97)

The prob. that the gaussian variable $x$ distributed as $G(x; \mu, \sigma)$ in in the range $(\mu - a, \mu + a)$ is

$$P(x \in (-a, a)) = erf\left(\frac{a}{\sigma\sqrt{2}}\right)$$  \hfill (1.98)

The Gaussian PDF is of such great importance not only because is the large-$\nu$ limit of the Poissonian but also because of the Central Limit Theorem:

Every random variable $X$ sum (or linear combination ) of many independent variables $x_i$ (i.e. $X = \sum_i x_i$) is distributed approximately as a Gaussian of mean $\sum_i \mu_i$ and variance $\sigma_X^2 = \sum_i \sigma_i^2$ in the limit $n \rightarrow \infty$ independently of the individual PDFs .

In practice, the CLT can be applied in many experimental situations in which the error is the sum of many independent causes: reading errors, instrumental noise, contaminations etc. In these cases, the measure can be assumed to be gaussian distributed.

Three important values of the cumulative function are

$$F(\mu + j\sigma) - F(\mu - j\sigma) = erf\left(\frac{j}{\sqrt{2}}\right) = 0.68, 0.95, 0.997$$  \hfill (1.99)

for $j = 1, 2, 3$: these give the prob. of finding $x$ at $j = 1, 2, 3\sigma$ from the mean $\mu$. Conventionally, errors are quoted at $1\sigma$ even for non-Gaussian distributions.
Figure 1.9: Comparing Poissonian and Gaussian PDFs for $\nu = 2$ (blue) and $\nu = 10$ (red).