11 Probability and Random Variables

*This introductory part is supposed to be **pre-requisite knowledge**. Students should be familiar with the general concept of probability and the basic probability theory results indicated here. The material reported in the following Sections 11.1 and 11.2 is **background material** and is intended as a support for those who do not remember the notions of probability theory provided in previous courses.*

11.1 What is "probability"? (background)

In the next few pages we will recall some of the foundational concepts of probability theory. An important disclaimer is that this must be understood as a refresher on the topic, and not as rigorous study material. For the latter, specific classes are available at Politecnico di Torino and countless books or plenty of on-line material can be easily found. Also, the approach followed here is empirical, whenever possible, rather than mathematical. Intuition is the substitute for rigor.

11.1.1 The key concepts

The key concepts in probability theory are that of the *random experiment* and of the *space of events, or sample space*.

The random experiment is a completely defined procedure which, every time it is executed, produces a result.

The space of events is the set of all the possible results of the random experiment.

For instance, a random experiment can be the rolling of a die. The corresponding space of events is a set that includes six elements, i.e., one for each face of the die.

Note that it is unimportant how the events are labeled. They can be labeled with numbers, or with colors, or with letters. If either letters or numbers are used, then the space of the events of rolling a die can be either:

$$\Omega = \{a, b, c, d, e, f\}$$

or

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Tossing a coin is another example of a random experiment, which has only two possible outcomes. The corresponding set can be labeled with numbers, but also with words, such as:

$$\Omega = \{0,1\} = \{head, tail\}$$

Both labeling strategies are fine, numbers or words. In the following, we will *always* use the labeling with numbers, which lend themselves better to dealing with more complex concepts related to probability.

11.1.2 Assigning a probability

We define N as the number of times an experiment is carried out. We then define $n_{\{i\}}$ as the number of times that the result $\{i\}$ presents itself. Then, we state that the "probability of the event $\{i\}$" is given by:

$$P\left(\{i\}\right) = \lim_{N\to\infty} \frac{n_{\{i\}}}{N}$$

Eq. 11-1

Note that our use of the symbol "lim" is not rigorous and in any case it does not coincide with that of calculus. It is possible to redefine the symbol "lim" in suitable mathematical terms for this context. On the other hand, we appeal to an intuitive "meaning", that is that $P\left(\{i\}\right)$ is the number which the ratio $n_{\{i\}} / N$ "tends to", with "increasing accuracy", as the value of N goes up.

The fact that such convergence does indeed take place, as the number N of repetitions of the experiment is increased, is stated by the so-called "Borel's law of large numbers". Note that this principle borders on the philosophical and in fact over the centuries various criticism has been thrown at it. For the interested reader, see:

http://en.wikipedia.org/wiki/Frequentist_probability

Nonetheless, it is a principle that has widespread, justified acceptance. In addition, it is a simple and intuitive way of defining "probability". Therefore, we adopt it here.

In certain cases, such as rolling a die, probability can also be inferred based on physical and logical arguments, such as evident symmetries. In other, more complex cases, this may not be possible and Borel's law may be the only practical way of deriving a probabilistic description of a certain random physical process.

Note that the quantity $P(\{i\})$, given its definition, can also be expected to *predict* how frequently a certain outcome of the experiment will present itself in the future. It is this predictive feature of probability theory that makes it one of the most important disciplines of the modern body of knowledge.

11.1.3 Sets of elementary events

Having embraced definition Eq. 11-1, then certain properties of the "probability" numbers follow suit.

Before listing them, we need to somewhat generalize the definition of an "event". In the case of rolling a die, the "elementary" outcomes of the random experiment are obviously one of the six faces pointing upward. However, one could also define a "composite" event as, for instance, "the result being an even number". This can be defined as a subset of $\Omega = \{1, 2, 3, 4, 5, 6\}$, and specifically:

$$A = \{2, 4, 6\}$$

Probabilities can be assigned to composite events as well, that is to subsets of Ω such as A, that are not just a single elementary event within Ω. Borel's law can be used for these events too. In the above example of $A = \{2, 4, 6\}$, it is:

$$P(A) = P(\{2,4,6\}) = \lim_{N\to\infty} \frac{n_{\{2,4,6\}}}{N}$$

where $n_{\{2,4,6\}}$ is the number of times either 2, 4 or 6 are observed as outcome of the random experiment. In general, a “probability” can ideally be defined in terms of Borel’s law for any possible subset of $A \subseteq \Omega$, through the obvious generalization:

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

Eq. 11-2

where n_A is the number of times the (possibly non-trivial) event $A \subseteq \Omega$ presents itself after the experiment.

Some of these events are special cases that can be immediately assigned a value. Specifically, $A = \Omega$ and $A = \Phi$, where $\Phi = \{\ \}$ is the empty set:

$$P(\Omega)=1 \qquad P(\Phi)=0$$

These two results are immediately justifiable in terms of Borel's law. In particular:

$$P(\Omega)=P(\{1,2,3,4,5,6\})=\lim_{N\to\infty}\frac{n_{\{1,2,3,4,5,6\}}}{N}=\frac{N}{N}=1$$

The result is obvious, noting that $n_{\{1,2,3,4,5,6\}}$ is the number of times the die shows either 1,2,3,4,5, or 6. Since this "event" occurs every time the die is rolled, that is: $n_{\{1,2,3,4,5,6\}}=N$, and the above result follows immediately. In fact, it is even true for any value of N, not necessarily for $N\to\infty$. Also:

$$P(\Phi) = P(\{\ \}) = \lim_{N \to \infty} \frac{n_{\{\ \}}}{N} = \frac{0}{N} = 0$$

This result too is found for any value of N, without the need for $N \to \infty$. The reason is that $n_{\{\ \}}$ is the number of times that, after rolling the die, none of the values 1,2,3,4,5, or 6 shows up. Since this is impossible by definition (the experiment must return one of those numbers), then this event never occurs and $n_{\{\ \}} = 0$.

Another important consequence of Borel's law is that probability is always a positive quantity:

$$P(A) \geq 0$$

This is also obvious from the definition Eq. 11-2, as both n_A and N are positive numbers. So in general, we can state that:

$$0 \le P(A) \le 1 \qquad \forall A \subseteq \Omega$$

11.1.4 Deriving probabilities for sets of events

As we said in the previous section, by the term "**elementary event**" we mean the direct outcome of the random experiment in its simplest form (rolling a die: a number between 1 and 6). Elementary events are all *mutually exclusive*, that is, it is not possible that the random experiment produces two or more of them at the same time. We also stated that non-elementary events, or **composite events**, are **sets of elementary events**.

While elementary events are mutually exclusive, composite events can be *non-exclusive*. To show this, let us consider for instance the following two composite events: "the result of rolling a die is an odd number", which corresponds to the set of elementary events: $A = \{1, 3, 5\}$; "the result of rolling a die is less than 4", which is the set $B = \{1, 2, 3\}$. Clearly, if the result of rolling the die is 3, both the composite events A and B occur at the same time, so A and B are *not* mutually exclusive.

In the following, we will assume that the probability of each "elementary event" in Ω has been found, either through Borel's law (experimentation) or through some physical argument (symmetries, etc.). We then show how to mathematically calculate the probability of a variety of composite events, without resorting again to Borel's law to estimate it.

11.1.4.1 Additivity

Given the composite event A, made up of a collection, or *union*, of *elementary events* a_k as follows:

$$A = \{a_1\} \cup \{a_2\} \cup \{a_3\} ... \cup \{a_K\} = \bigcup_{k=1}^{K} \{a_K\} = \{a_1, a_2, a_3, ..., a_K\}$$

$$a_1, a_2, a_3, ..., a_K \in \Omega$$

then the probability of A is:

$$P(A) = \sum_{k=1}^{K} P(\{a_k\})$$
$$= P(\{a_1\}) + P(\{a_2\}) + P(\{a_3\}) + \ldots + P(\{a_K\})$$

Eq. 11-3

For instance, looking at the "roulette" casino game, the space of events is $\Omega = \{0,1,2,3,\ldots,36\}$. The probability of each elementary outcome is identical and is 1/37. If we are interested in the probability of any of the so-called "orphans" coming out, that is:

$$A = \{1,6,9,14,17,20,31,34\}$$

then this is quite simply:

$$P(A) = P(\{1\}) + P(\{6\}) + P(\{9\}) + P(\{14\}) + P(\{17\}) +$$
$$+ P(\{20\}) + P(\{31\}) + P(\{34\}) = \frac{8}{37} \approx 0.2162$$

The additivity rule for elementary events can be easily proved based on Borel's law. From Eq. 11-3 we write:

$$\sum_{k=1}^{K} P(\{a_k\}) = \sum_{k=1}^{K} \lim_{N\to\infty} \frac{n_{\{a_k\}}}{N} = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{K} n_{\{a_k\}}$$

Eq. 11-4

But

$$\sum_{k=1}^{K} n_{\{a_k\}} = n_{\{a_1\}} + n_{\{a_2\}} + \ldots + n_{\{a_K\}} = n_{\{a_1, a_2, \ldots, a_K\}} = n_A$$

Eq. 11-5

.

So, substituting Eq. 11-5 into Eq. 11-4, we get:

$$\sum_{k=1}^{K} P(\{a_k\}) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{K} n_{\{a_k\}}$$

$$= \lim_{N\to\infty} \frac{n_{\{a_1, a_2, \ldots, a_K\}}}{N} = \lim_{N\to\infty} \frac{n_A}{N} = P(A)$$

which proves the additivity rule, Eq. 11-3.

The additivity rule of probability works also for the union of composite events, provided that the two composite events are mutually exclusive. Note that two composite events are mutually exclusive if and only if they are represented by *disjoint sets*, that is, sets with empty intersection.

Given:

$$A = \{a_1, a_2, a_3, \ldots, a_K\}$$

$$B = \{b_1, b_2, b_3, \ldots, b_M\}$$

then:

$$P\left(A \bigcup B\right) = P\left(A\right) + P\left(B\right)$$

provided that

$$A \bigcap B = \varnothing$$

that is:

$$a_k \neq b_m \quad k = 1,\ldots,K \quad m = 1,\ldots,M$$

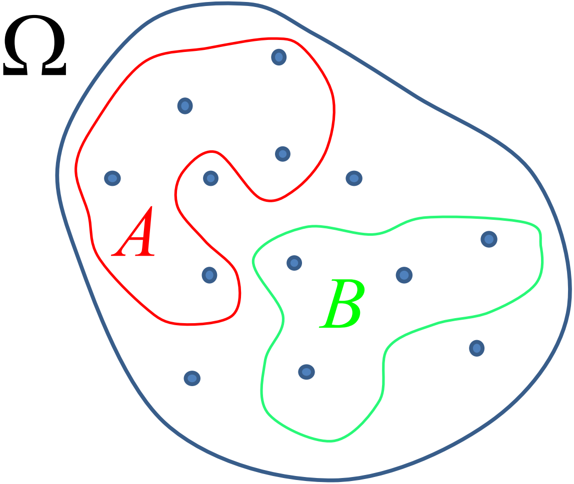


Fig. 11-1: The sets A and B are "disjoint", that is: $A \cap B = \varnothing$. They represent "mutually exclusive" composite events.

In other words, the additivity rule holds as long as all the elementary events present in A are not present in B and vice versa.

11.1.4.2 Non-disjoint events

It is possible that, given:

$$A = \left\{ a_1, a_2, a_3, ..., a_K \right\}$$

$$B = \left\{ b_1, b_2, b_3, ..., b_M \right\}$$

then

$$a_k = b_m \quad \text{for some } k \text{ and } m$$

Put it differently, the composite events A and B are non-disjoint, or non-mutually exclusive, i.e., the have a non-empty intersection:

$$A \bigcap B = C \neq \varnothing$$

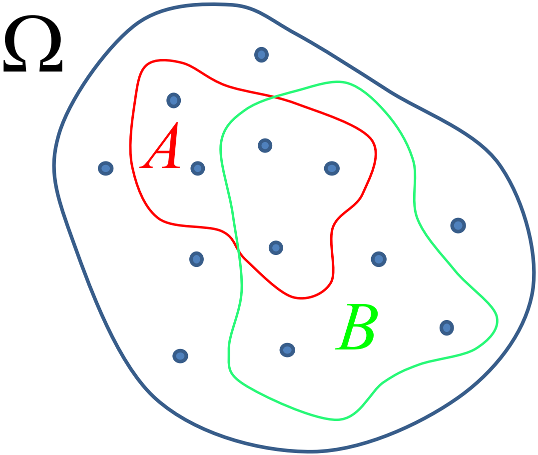


Fig. 11-2: The composite sets A and B are "non-disjoint", or "non-mutually exclusive", that is: $A \cap B \neq \varnothing$

When this is the case, the two events are actually non-mutually exclusive. If so, then $P(A \cup B) \neq P(A)+P(B)$. However, a simple "correction" can be used to get a general formula that can deal with non-exclusivity. In fact, this correction generalizes the rule to all possible cases.

Given two events A and B, either elementary or composite, disjoint or non-disjoint, then:

$$P(A \cup B)=P(A)+P(B)-P(C)$$
$$\text{where } C=A \cap B$$

This is easily proved as follows. If A and B are disjoint, then $C=\varnothing$. Since $P(\varnothing)=0$, in this case we get the previous result $P(A \cup B)=P(A)+P(B)$, as it should be. So, the general formula includes the case of disjoint sets correctly.

If instead the sets are not disjoint, then this means that some of the elementary events in A are also present in B. In fact, they are precisely the elementary events collected in C. If probabilities were simply added, as $P(A)+P(B)$, then the elementary events that are both present in A and B would contribute their individual probability *twice*, once in $P(A)$ and once in $P(B)$, as if the probability of these elementary events would double. This is not correct, because the probability of any individual elementary event cannot change. Therefore, it is necessary to remove this *doubling* of probability by subtracting it *once*. These elementary events are all those of C, and therefore the correct result is $P(A)+P(B)-P(C)$.

Notice that if one executes first the set operation: $A \cup B = D$, and then finds the probability $P(D)$, then there is no need of any correction, because the "union" of two sets does not duplicate the elements. In other words, if an element is both in A and in B, it appears only once in $D = A \cup B$, so its probability is correctly counted once in $P(D)$.

11.1.4.3 The conditional probability formula

The *"conditional probability" definition formula* recites:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Eq. 11-6

The quantity $P(A \mid B)$ can be read as the probability of the event A occurring given that the event B has occurred. Equivalently, and perhaps more clearly, it can be viewed as **the probability of the event A occurring in a space of events that has been reduced to just B.**

Two events A and B are said to be **statistically independent** if and only if:

$$P(A|B) = P(A)$$

Eq. 11-7

Condition Eq. 11-7 can be re-written as:

$$P_B(A) = P_\Omega(A)$$

Eq. 11-8

where $P_\Omega(A)$ (or simply $P(A)$) is the probability of A in the original space of events Ω, whereas $P_B(A)$ is the probability of A over the reduced space of events B.

Note that sometimes the concept of statistical independence is mistaken for the concept of mutual exclusivity or disjointness. _*The two concepts are very different and should not be confused*_.

If A and B are statistically independent then, combining Eq. 11-6 and Eq. 11-7, the ***intersection probability formula*** can be written as:

$$P(A \cap B) = P(A)P(B)$$

Eq. 11-9

If A and B are not statistically independent then, according to Eq. 11-6, the intersection probability formula[1] becomes:

$$P(A \cap B) = P(A \mid B)P(B) = P(B \mid A)P(A)$$

Eq. 11-10

---

[1] The relationship $P(A \mid B)P(B) = P(B \mid A)P(A)$ is very famous and is called Bayes' rule, or Bayes' theorem, and is one of the pillars of statistical data analysis. For more details, you can read https://en.wikipedia.org/wiki/Bayes%27_theorem

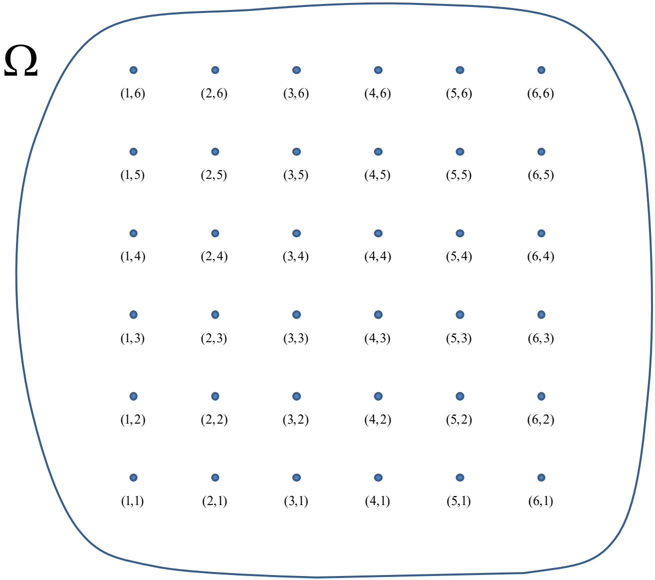
Note that, as shown, the role of A and B can be swapped in the right-hand side.

11.1.4.4 Examples of intersection probabilities

Some explanatory examples of the intersection probability formulas Eq. 11-9 and Eq. 11-10 can be easily obtained by looking at the random experiment consisting of simultaneously rolling two dice, which we call dice 1 and dice 2. The elementary event which is the outcome of each random experiment is a pair of numbers:

$$(m,n)$$
$$m \in \{1,2,3,4,5,6\}, \quad n \in \{1,2,3,4,5,6\}$$

We can also say that the space of events Ω is the following:



We assume that the dice are perfect cubes of uniform mass density and hence their center of mass is located at the geometric center of the cubes themselves. If so, there is no reason whatsoever a certain face should end up showing on top more frequently than any other die face. Also, there is no reason why the outcome of rolling one die should influence the outcome of rolling the other die. This means that the probability of anyone of the elementary outcomes of the random experiment in the space of events Ω is the exact same. The total number of events is 36 and hence the probability of anyone of the elementary events is:

$$P\{(m,n)\} = \frac{1}{36}$$

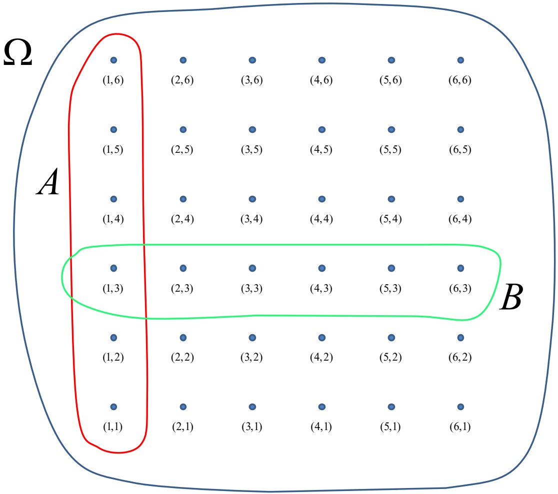
$$\forall m \in \{1,2,3,4,5,6\}, \quad \forall n \in \{1,2,3,4,5,6\}$$

We now define two composite events:

$$A = \{\text{first die is 1, second die any}\} = \{m = 1\}$$

$$B = \{\text{second die is 3, first die any}\} = \{n = 3\}$$

Pictorially:



The two composite events A and B are not mutually exclusive. **Are they statistically independent?** To find out, we can perform the check of either Eq. 11-7 or Eq. 11-8. We start with Eq. 11-7. Is it true that $P(A \mid B) = P(A)$?

To calculate the left-hand side we use the conditional probability formula Eq. 11-6:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} =$$

$$= \frac{P\{(1,3)\}}{P\{(1,3),(2,3),(3,3),(4,3),(5,3),(6,3)\}} = \frac{1/36}{6/36} = \frac{1}{6}$$

As for the right-hand side, we get:

$$P(A) = P\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6)\} = 6/36 = \frac{1}{6}$$

Eq. 11-11

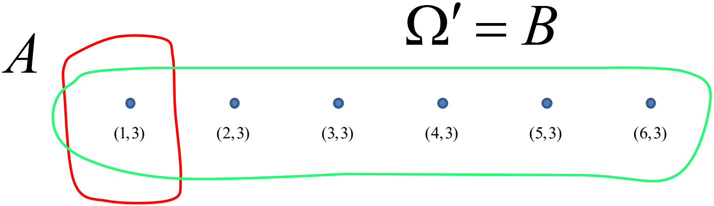
Therefore, indeed:

$$P(A|B) = P(A) = \frac{1}{6}$$

which proves that A *and* B *are statistically independent events.*

The same result could be obtained by checking if Eq. 11-8 was verified, that is if: $P_B(A) = P_\Omega(A)$.

The right-hand side is the same Eq. 11-11. The left-hand side requires that a new space of events, say Ω', is defined, which coincides with B. If so, we pictorially get:



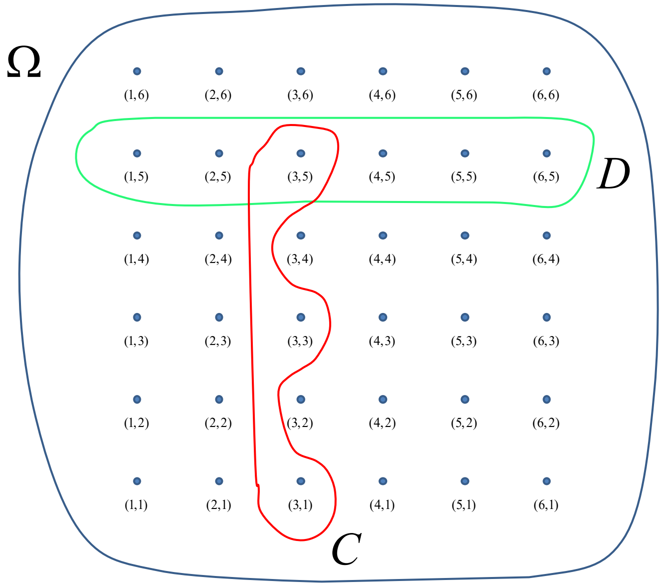
Then we have to calculate the probability of A occurring in $\Omega' = B$. Clearly the probability is $1/6$, so we indeed verify that:

$$P_B\left(A\right) = P_\Omega\left(A\right) = 1/6$$

The next example will show a case in which there is not statistical independence between two composite events C, D. They are defined as:

$$C = \left\{\text{first die is 3, second die is odd}\right\} = \left\{\left(3,1\right),\left(3,3\right),\left(3,5\right)\right\}$$
$$D = \left\{\text{first die any, second die is 5}\right\} = \left\{\left(1,5\right),\left(2,5\right),\left(3,5\right),\left(4,5\right),\left(5,5\right)\right\}$$

Pictorially:



Again, we check for statistical independence by computing the two sides of the condition Eq. 11-7: $P(C\,|\,D) = P(C)$.

The left-hand side can be calculated once more with the general formula:

$$P(C \mid D) = \frac{P(C \cap D)}{P(D)} =$$

$$= \frac{P\{(3,5)\}}{P\{(1,5),(2,5),(3,5),(4,5),(5,5),(6,5)\}} = \frac{1/36}{6/36} = \frac{1}{6}$$

As for the right-hand side, we get:

$$P(C) = P\{(3,1),(3,3),(3,5)\} = 3/36 = \frac{1}{12}$$

Now we see that:

$$P(C \mid D) = \frac{1}{6} \quad \neq \quad P(C) = \frac{1}{12}$$

which tells us that the composite events C, D, are ***not*** statistically independent.

As a further check, we can ascertain whether Eq. 11-9 is verified or not. Since C, D, are not statistically independent, it should not hold. In fact it does not:

$$P(C \cap D) = P\{(3,5)\} = \frac{1}{36} \neq P(C)P(D) = \frac{1}{6}\frac{1}{12} = \frac{1}{72}$$

$\boxed{\text{On your own:}}$ perform the check using the condition $P_D(C) = P_\Omega(C)$ and verify that it is not met either.

$\boxed{\text{On your own:}}$ why is it, that C, D, are not statistically independent? Can you find an explanation?

11.1.4.5 independence and mutual exclusivity

As mentioned, statistical independence and mutual exclusivity are different concepts. In fact, they are <u>incompatible conditions</u>:

given an event space Ω *and two non-empty events in* Ω*:* $A \neq \varnothing$ *and* $B \neq \varnothing$*, with non-zero probability* $P(A) \neq 0$ *and* $P(B) \neq 0$*; then, <u>if</u>* A *<u>and</u>* B *<u>are mutually exclusive (that is</u>* $A \cap B = \varnothing$*<u>), then</u>* A *<u>and</u>* B *<u>cannot be statistically independent</u>.*

This is shown as follows. For A and B to be independent it must be: $P(A|B) = P(A)$. However:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = 0$$

because the intersection of the two events is empty: $P(A \cap B) = P(\varnothing) = 0$. Hence, it is impossible that $P(A|B) = P(A)$, since $P(A) \neq 0$.

11.2 Random variables (background)

Random variables (RVs) are a special case of a random experiment. Specifically, it is assumed that the space of events is not just a set of numbers, but it is the whole of $\mathbb{R}$. In other words, it is assumed that the result of the random experiment can potentially be any real number.

RV are typically denoted with a Greek letter, such as ξ, or η, or ρ. Note that these symbols *do not denote just numbers*: they represent a random experiment, a set of resulting events (any number over the whole of $\mathbb{R}$) and some probabilistic description of the outcome of the experiment.

The fact that the possible results are infinite (as many as the real numbers), and not even countable, has consequences. In particular, the probabilistic description of a random variable cannot be provided by simply assigning a positive probability to each possible result. If the possible results are infinite, even assigning a very small finite (but non-vanishing) probability to each one of them would make

$P(\Omega) = P(\mathbb{R}) = \infty$. But this is absurd, since, as we know, $P(\Omega) = 1$. Also, any direct link to Borel's law would fail to be established, raising problems of interpretation of what these statistical objects are.

11.2.1 the probability histogram

A different approach must then be used. One can for instance think of dividing the real axis into "bins", that is, intervals of values, which represent sets of possible results of the random experiment. For instance, one can identify the *i*-th "bin" as centered about a value $x_i = i \cdot \Delta x$, and being Δx wide. This way, the *i*-th "bin", which we could call Δx_i, would correspond to the interval:

$$\Delta x_i = \left[x_i - \frac{\Delta x}{2}, x_i + \frac{\Delta x}{2} \right]$$

It is then obvious that, if we allow *i* to span all integers, then the full set $\{\Delta x_i\}_{i=-\infty}^{\infty}$

actually covers the whole of $\mathbb{R}$, i.e., the whole space of events: $\left\{\Delta x_i\right\}_{i=-\infty}^{\infty} = \mathbb{R} = \Omega$

.

Then, instead of asking ourselves what the probability is of each real number being the outcome of the random experiment, we can ask ourselves what the probability is that the result of the random experiment ends up within any one bin Δx_i.

These probabilities have the advantage that they can then be easily related to Borel's law. Specifically, we can define the probability that the random variable ξ takes a value within the *i*-th bin as:

$$P\left(\left\{\xi \in \Delta x_i\right\}\right) = \lim_{N\to\infty} \frac{n_{\{\xi\in\Delta x_i\}}}{N}$$

Eq. 11-12

where $n_{\{\xi\in\Delta x_i\}}$ is the number of times this happens, out of N experiments.

The set of values given by Eq. 11-12 can be associated to the bars of a histogram along x and represented as such. This "probability histogram" constitutes one possible way of characterizing a RV.

Notice that, assuming that indeed $N \to \infty$ and the left-hand side of Eq. 11-12 converges to its exact value, then we can easily show that

the sum of the values of all the "histogram bars" is one.

In fact, we can write:

$$\sum_{i=-\infty}^{\infty} P\left(\left\{\xi \in \Delta x_i\right\}\right) = P\left(\bigcup_{i=-\infty}^{\infty} \left\{\xi \in \Delta x_i\right\}\right) = P\left(\mathbb{R}\right) = P\left(\Omega\right) = 1$$

Eq. 11-13

where we have used in reverse the rule of the probability of the union of disjoint events, thanks to the fact that the events $\left\{\xi \in \Delta x_i\right\}$ are all disjoint (because the Δx_i 's are too).

Quite interestingly, *if* N *was finite*, as in the case of a practical experimental characterization of the RV, the values of the bars would be affected by some error but:

the sum of the values of all the bars would still be one, even for a finite number of trials N.

We prove this in the following. The sum of all the histogram bars is:

$$\sum_{i=-\infty}^{\infty} P\left(\left\{\xi \in \Delta x_i\right\}\right) = \sum_{i=-\infty}^{\infty} \lim_{N\to\infty} \frac{n_{\{\xi\in\Delta x_i\}}}{N} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=-\infty}^{\infty} n_{\{\xi\in\Delta x_i\}}$$

Eq. 11-14

But of course: $\sum_{i=-\infty}^{\infty} n_{\{\xi\in\Delta x_i\}} = n_{\{\xi\in\mathbb{R}\}} = N$, so that:

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=-\infty}^{\infty} n_{\{\xi\in\Delta x_i\}} = \lim_{N\to\infty}\frac{1}{N}\cdot N = 1$$

Eq. 11-15

Even removing the limit, the result would always be one, since of course $\frac{1}{N}\cdot N = 1$ independently of how big N is. Therefore, even in the case of an *estimation* of probabilities, based on a finite number of random experiments, we would still have a histogram with values summing to one, which is an interesting circumstance and makes the estimation well-behaved.

11.2.2 from histogram to "pdf"

For various reasons, probability histograms may appear as a sort of inelegant and not entirely satisfactory RV characterization. One might wonder whether it would be possible to replace a histogram with some sort of smooth curve over x which might provide the same, if not better, information.

One idea would be to try to shrink the bin size Δx towards zero. However, as long as the bin size Δx is perhaps small, but finite, then the approach, at least *ideally*, poses no theoretical problems. On the other hand, if the size of Δx is really made to go to zero, then the interval Δx_i tends to become the single number x_i and the problem of dealing with the probability of a single number being the result of the random experiment, among an infinity of possible results, arises again. Essentially, once again, such probability would typically be zero. One would end up with an all-zero curve, which has no meaning.

To solve this problem, people have resorted to a very widely used approach, which is that of using *densities*.

An example may clarify the approach. For instance, one might want to characterize the mass of a straight steel bar in its longitudinal l direction, whose cross-section diameter is irregular. Ideally, he/she would have to provide the mass at each point along l, but this technically results in zero mass at all points, since a single point along l carries no mass (because it has no length). The problem is solved by defining a *mass density* along l, specifically $\mu(l)=\frac{dm}{dl}$, which allows to characterize the steel bar mass point-by-point, even though indeed a single longitudinal point of the steel bar along l has no mass.

The same can be done to characterize the probability of a random variable along x. The probability of each single point x is typically zero, but the probability density along x may not be zero. We define the *density of probability at the point* x_i, which is typically denoted as $f_\xi(x_i)$, as:

$$f_\xi(x_i)=\lim_{\Delta x\to 0}\frac{P(\{\xi\in\Delta x_i\})}{\Delta x}$$

Eq. 11-16

Since the location of the point x_i in the left-hand side above is arbitrary, we can remove the index i and simply write:

$$f_\xi(x) = \lim_{\Delta x \to 0} \frac{P\left(\left\{\xi \in \left[x - \Delta x/2, x + \Delta x/2\right]\right\}\right)}{\Delta x}$$

Note that such definition can be viewed as a derivative of probability vs. x, that is:

$$f_\xi(x) = \frac{dP}{dx}$$

much as we could view: $\mu(l) = \frac{dm}{dl}$. If we want to know how much probability is present within a certain interval, we then only need to do integrate over the interval:

$$P\left(\left\{\xi \in [a,b]\right\}\right) = \int_a^b f_\xi(x)\,dx$$

Eq. 11-17

in the same way as to calculate the mass of a bar between the longitudinal points a and b we need to do:

$$m\left([a,b]\right) = \int_a^b \mu(l)\,dl$$

It is also easy to prove that the following properties hold:

$$f_\xi(x) \ge 0$$

Eq. 11-18

$$\int_{-\infty}^{\infty} f_\xi(x)\,dx = P\left(\left\{\xi \in [-\infty,\infty]\right\}\right) = P(\mathbb{R}) = P(\Omega) = 1$$

Eq. 11-19

It turns out that the **probability density function (pdf)** truly is a complete and full description of the RV ξ.

End of background material (sects. 11.1 and 11.2)

11.3 Histograms and approximate pdf

As we saw in the previous sections, a random variable ξ is completely and fully characterized through its **probability density function (pdf)**, $f_\xi(x)$. So, when we want to describe a certain random event in terms of a random variable, what we need to do is essentially to find out what the pdf of that random variable is.

Probability density functions for many events can often be found a priori, based on analytical and physical arguments. In those cases, rigorous derivations based on physical laws lead to the pdf of the random variable. For instance, many fundamental physical effects produce random quantities whose pdf can be predicted to be *Gaussian*. Random variables whose pdf is Gaussian are very important in many contexts and we will look at the Gaussian pdf in detail later on.

In some cases the random experiment is too complicated for the pdf to be calculated a priori and can only be estimated *experimentally*. In these cases, it is possible to use a histogram analysis of the values generated by the random variable through repeated experiments, to produce an *approximation* of $f_\xi(x)$.

To do so, we first focus on certain specific outcomes of the random experiment, that is, on certain specific values of x, say, x_i. A reasonable choice is to choose such points so that the are equally spaced, that is:

$$x_{i+1} - x_i = \Delta x \quad , \quad \forall i$$

Then, a possible approximation of $f_\xi(x_i)$ is:

$$f_\xi(x_i) \approx \frac{n_{\{\xi \in \mathbf{I}_i\}}}{N \Delta x}$$

Eq. 11-20

where $n_{\{\xi \in \mathbf{I}_i\}}$ is the number of times the random experiment generates a result that is comprised within the interval $\mathbf{I}_i = \left[x_i - \frac{\Delta x}{2}, x_i + \frac{\Delta x}{2} \right]$, out of the total N times the random experiment is carried out.

This is an approximation of $f_\xi(x_i)$ and not an exact value because of two reasons.

One reason is that each ratio $n_{\{\xi\in\mathbf{I}_i\}}/(N\Delta x)$ fluctuates randomly as N is increased. It would ideally converge to a stable value only if an infinity of random experiments ($N\to\infty$) was carried out. Also, the accuracy of the approximation tends to improve as N goes up.

The other reason is that Δx should be vanishingly small, in order for $n_{\{\xi\in\mathbf{I}_i\}}/(N\Delta x)$ to precisely converge to $f_\xi(x_i)$ as N goes up.

In practice, one would generate as many values as possible and use as small as possible values of Δx, keeping in mind that if Δx is reduced, then N has to be increased, otherwise each $n_{\{\xi\in\mathbf{I}_i\}}$ may become too small and be affected by too much random uncertainty.

The number of observed results in a bin $n_{\{\xi \in \mathbf{I}_i\}}$ that guarantees "sufficient" reliability to that bin height estimate can be characterized rigorously, but <u>we will not deal with this aspect in detail here</u>. Also, the required level of reliability may vary depending on applications. Certainly, a few results per bin is likely to be insufficient for most applications. Instead, several tens or hundreds are likely to produce rather good estimates.

If we want to write down the the generated histogram as a function of the continuous variable $x \in \mathbb{R}$, this can be obtained by assigning to each interval $\mathbf{I}_i$ a rectangular function of x with height $n_{\{\xi \in \mathbf{I}_i\}} / (N\Delta x)$:

$$f_\xi(x) \approx f_{\xi,\text{app}}(x) = \sum_{i=-\infty}^{\infty} \frac{n_{\{\xi \in \Delta x_i\}}}{N\Delta x} \cdot \Pi_{\Delta x}(x - x_i)$$

Eq. 11-21

This approximation $f_{\xi,\text{app}}(x)$ has the nice feature that it *always integrates to 1*, even though it is just an approximation of the actual pdf $f_{\xi}(x)$, for any number of trials N and for any size of the "bins" Δx. That is

$$\int_{-\infty}^{\infty} f_{\xi,\text{app}}(x)\,dx = 1 \quad , \quad \forall N, \quad \forall \Delta x$$

$\boxed{\text{On your own:}}$ prove that the integral of $f_{\xi,\text{app}}(x)$ over the whole of $\mathbb{R}$ is always one. Hint: direct integration leads immediately to the result.

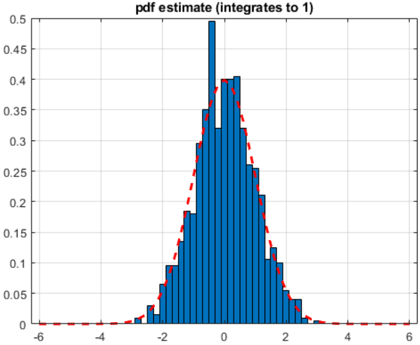
This histogram method of estimating a pdf is very effective when a stochastic problem can be dealt with in the context of modern numerical simulation tools that allow to perform millions or billions of virtual "experiments" in a short time.

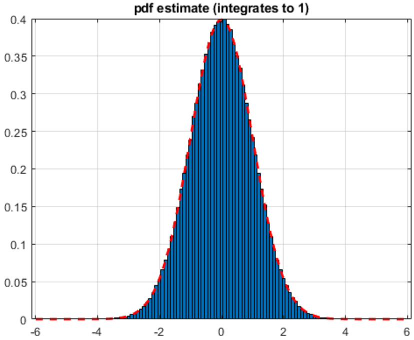
Apart from the practical aspects, the histogram approximation provides good intuitive insight into what a pdf actually means.

If you are interested in reading more about this method and Borel's law of large numbers, you can start here https://en.wikipedia.org/wiki/Law_of_large_numbers

Incidentally, under the section title "Consequences" the pdf approximation method through a histogram is reported.

Here below two examples. In both cases the actual pdf is Gaussian with variance 1 and mean value 0. In the left plot, $N = 1000$ and $\Delta x = 0.2$. In the right plot $N = 1000000$ and $\Delta x = 0.1$. The dashed red curve is the exact pdf.





11.3.1 Discrete Random Variables

Certain random experiments, like rolling a die, produce a *countable* set of results. As we have seen at the beginning of this section, such events can be described relatively easily by labeling the elementary events in Ω and then quoting each one of them as to their individual probability. The machinery of random variables (RV's) is not strictly needed. Nonetheless, it would be nice to try to unify under a single description both countable and uncountable spaces of events.

This is in general possible: when the random experiment produces a countable set of discrete results $\Omega = \{x_k\}_{k=1}^{K}$, where K could also be infinity, the resulting pdf is:

$$f_\xi(x) = \sum_{k=0}^{K} P(\{\xi = x_k\})\,\delta(x - x_k)$$

Such pdf actually complies with all the theoretical requirements of any pdf.

Referring to the die example, the RV can of course take on only six values, for instance:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

(though any other choice of 6 distinct numbers would also do). Then, the pdf would be:

$$f_{\xi}(x) = \sum_{k=1}^{6} \frac{1}{6} \delta(x-k)$$

11.4 The "Expectation Operator"

From now on, the material has been presented in class and discussed.

11.4.1 Definition: averaging over one random variable

The "expectation operator", or "expected value" operator is one of the most important operators used in statistics. We first provide its definition in connection with a single random variable. We'll then extend it to any number of variables.

We define the *expected value operator* or simply *expectation* operator as follows:

$$E_\xi\{g(\xi)\} \triangleq \int_{-\infty}^{+\infty} g(x) f_\xi(x) dx$$

Eq. 11-22

where $g(x)$ is any real or complex function of the independent variable $x \in \mathbb{R}$.

Note that depending on the nature of $g(x)$, the integral may not converge. However, in this course, this situation does not arise.

11.4.1.1 Linearity

The expectation operator is a linear operator. Given a linear combination of functions of the random variable ξ, we have:

$$E_\xi\left\{\sum_{n=1}^{N} a_n\, g_n(\xi)\right\} = \sum_{n=1}^{N} a_n\, E_\xi\{g_n(\xi)\}$$

The proof is straightforward, since the expectation operator is essentially an integral which is, by itself, a linear operator.

11.4.1.2 Expectation of an independent quantity

The expectation value of any quantity that is not related to the random variable ξ coincides with the quantity itself:

$$E_{\xi}\{g(z)\} = g(z)$$

Eq. 11-23

Note that this is true even if the argument depends on a random variable, provided that such random variable is statistically independent of the one over which the operator is taking the expectation. For instance, given a random variable η that is statistically independent of ξ, then:

$$E_{\xi}\{g(\eta)\} = g(\eta)$$

Note that if η was instead statistically dependent on ξ, then the result of $E_{\xi}\{g(\eta)\}$ would be different and would require the knowledge of the relationship between η and ξ. This topic will be dealt with later on.

As a special case, the expectation of any constant is clearly the constant itself (prove it on your own):

$$E_{\xi}\{a\} = a$$

11.4.1.3 Transfer of expectation

This is a very important and powerful property, which is often used in practical calculations.

Let us assume that we need to evaluate $E_{\xi}\{g(\xi)\}$. We also assume that ξ can be obtained as a deterministic function of another random variable η, so that: $\xi = q(\eta)$.

Then:

$$E_{\xi}\{g(\xi)\} = E_{\eta}\{g(q(\eta))\}$$

Eq. 11-24

Note that the expectation in the right-hand side is now taken over η. We therefore have:

$$E_{\xi}\{g(\xi)\} = E_{\eta}\{g(q(\eta))\} = \int_{-\infty}^{+\infty} g(q(z)) f_{\eta}(z) dz$$

Eq. 11-25

The proof of this property is easy in selected special cases, such as when the function $q(z)$ is monotonic and increasing or monotonic and decreasing. In the general case, however, it is difficult, so we omit it.

11.4.2 Non-central moments

Non-central moments are expectations of functions of a random variable of the form: $g(\xi) = \xi^n$. A non-central moment of order n of the random variable ξ is then:

$$E_\xi\{\xi^n\} = \int_{-\infty}^{+\infty} x^n f_\xi(x)dx$$

Eq. 11-26

The most important non-central moments are found for $n = 1, 2$. They have specific names. The moment with $n = 1$ is called *mean value of* ξ, or *statistical average of* ξ, or *expected value of* ξ, and is typically labeled with the symbol μ_ξ:

$$\mu_\xi = E\{\xi\} = \int_{-\infty}^{+\infty} x\, f_\xi(x)dx$$

Eq. 11-27

For $n = 2$ it is called *mean square value of* ξ*:*

$$E\{\xi^2\} = \mathrm{MSV}\left\{\xi\right\} = \int_{-\infty}^{+\infty} x^2\, f_\xi(x)dx$$

Eq. 11-28

A quantity derived from the mean square value of ξ is universally designated with the acronym RMS ("root-mean-square") and is simply:

$$\mathrm{RMS}\left\{\xi\right\} = \sqrt{E\{\xi^2\}}\,.$$

11.4.2.1 Optional: moments and Borel's law

An interesting question is whether a moment can be estimated using Borel's law of large numbers and whether the result would somehow agree with the integral definitions above. The answer is affirmative. For instance, it can be shown that:

$$\mu_\xi = \lim_{N\to\infty} \frac{x_1 + x_2 + \ldots + x_N}{N}$$

Eq. 11-29

where $x_1, x_2, \ldots, x_N$ are the results of N random experiments, that is, they are the values taken on by the RV ξ over N executions of the random experiment.

Eq. 11-29 actually coincides with Eq. 11-27 in the indicated limit of the number of trials N going to infinity. This is in general true for all expectations, in the sense that given a generic expectation $E_\xi\{g(\xi)\}$, then:

$$E_\xi\{g(\xi)\} = \lim_{N\to\infty} \frac{g(x_1)+g(x_2)+\ldots+g(x_N)}{N}$$

(end of optional)

11.4.2.2 Central moments

Central moments are expectations of functions of a random variable, of the form: $g(\xi)=(\xi-\mu_\xi)^n$. In other words, central moments are computed "about the mean value". They provide information regarding how a random variable departs from its mean value.

The central moment of order n of the random variable ξ is then:

$$E\left\{(\xi-\mu_\xi)^n\right\} = \int_{-\infty}^{+\infty} (x-\mu_\xi)^n f_\xi(x)dx$$

Eq. 11-30

By far the most important central moment is that of order 2, which is called *variance* and is typically labeled with the symbol σ^2:

$$\sigma_\xi^2 \triangleq E\left\{\left(\xi - \mu_\xi\right)^2\right\} = \int_{-\infty}^{+\infty}\left(x - \mu_\xi\right)^2 f_\xi(x)dx$$

Eq. 11-31

Note that a fully equivalent definition of the variance is:

$$\sigma_\xi^2 \triangleq E\left\{\xi^2\right\} - \mu_\xi^2$$

Eq. 11-32

On your own, prove that the two definitions are identical.

As for the mean square value, here too a much-used quantity is derived from the variance by taking its square root. Such quantity is called *standard deviation* and is typically labeled with the symbol σ:

$$\sigma_\xi = \sqrt{\sigma_\xi^2}$$

Eq. 11-33

11.4.2.3 Example: mean and variance of a uniformly-distributed RV

We assume a completely generic uniform distribution, such that ξ can take values in the generic interval $[a,b]$, then:

$$f_\xi(x) = \frac{1}{b-a}\Pi_{b-a}\left(x - \frac{a+b}{2}\right) \quad \Rightarrow$$

$$\Rightarrow \quad \mu_\xi = E\{\,\xi\} = \int_{-\infty}^{+\infty} x \cdot f_\xi(x)dx = \int_{-\infty}^{+\infty} x \cdot \frac{1}{b-a}\Pi_{b-a}\left(x - \frac{a+b}{2}\right)dx$$

$$= \frac{1}{b-a}\int_a^b x\,dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b \frac{1}{b-a}\left(\frac{b^2-a^2}{2}\right) = \frac{a+b}{2}$$

The statistical average coincides with the arithmetic average of the extremes of the distribution. This clearly happens because every value between the two extremes has the same statistical weight (or “probability density”) in the mean value calculation. As for the variance:

$$
\begin{aligned}
\sigma_\xi^2 &= E_\xi\left\{\left(\xi-\mu_\xi\right)^2\right\} = \int_{-\infty}^{+\infty}\left(x-\mu_\xi\right)^2 f_\xi(x)dx = \\
&= \frac{1}{b-a}\int_a^b\left(x-\mu_\xi\right)^2 dx = \frac{1}{b-a}\int_a^b\left(x^2-2x\mu_\xi+\mu_\xi^2\right)dx = \\
&= \frac{1}{b-a}\left[\int_a^b x^2dx-2\mu_\xi\int_a^b xdx+\mu_\xi^2\int_a^b 1(x)dx\right] = \\
&= \frac{1}{b-a}\left[\frac{x^3}{3}-2\mu_\xi\frac{x^2}{2}+\mu_\xi^2 x\right]_a^b = \\
&= \frac{1}{b-a}\left[\frac{b^3-a^3}{3}-2\mu_\xi\frac{b^2-a^2}{2}+\mu_\xi^2\left(b-a\right)\right] = \\
&= \frac{a^2+ab+b^2}{3}-\mu_\xi\left(a+b\right)+\mu_\xi^2 = \\
&= \frac{a^2+ab+b^2}{3}-\frac{\left(a+b\right)^2}{2}+\frac{\left(a+b\right)^2}{4} = \\
&= \frac{4\left(a^2+ab+b^2\right)-6\left(a+b\right)^2+3\left(a+b\right)^2}{12} = \\
&= \frac{a^2-2ab+b^2}{12} = \frac{\left(b-a\right)^2}{12};
\end{aligned}
$$

$\boxed{\text{The}}$ The calculation can be re-done using the alternative formula for the variance: $\sigma_\xi^2 = E_\xi\left\{\ \xi^2\right\} - \mu_\xi^2$. We already know μ_ξ, so we need to find:

$$E_\xi\left\{\ \xi^2\right\} = \int_{-\infty}^{+\infty} x^2 f_\xi(x)dx = \int_{-\infty}^{+\infty} x^2 \frac{1}{b-a}\Pi_{b-a}\left(x - \frac{a+b}{2}\right)dx =$$

$$= \frac{1}{b-a}\int_a^b x^2 dx = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b = \frac{1}{b-a}\frac{b^3-a^3}{3}$$

$$= \frac{1}{b-a}\frac{(b-a)\cdot\left(a^2+ab+b^2\right)}{3} = \frac{a^2+ab+b^2}{3}$$

We then put the two results together:

$$
\begin{aligned}
\sigma_\xi^2 &= E_\xi\left\{\xi^2\right\} - \mu_\xi^2 \\
&= \frac{a^2+ab+b^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{a^2+ab+b^2}{3} - \frac{a^2+2ab+b^2}{4} \\
&= \frac{4\left(a^2+ab+b^2\right)-3\left(a^2+2ab+b^2\right)}{12} = \\
&= \frac{4a^2+4ab+4b^2-3a^2-2ab-3b^2}{12} = \frac{a^2-2ab+b^2}{12} = \frac{\left(b-a\right)^2}{12};
\end{aligned}
$$

It is immediately evident that the variance is totally independent of the mean value μ_ξ. It only depends on the extension of the pdf. In fact, in this case:

$$
\operatorname{ext}\left\{f_\xi(x)\right\} = \left[a,b\right]
$$

Calling the "width" of the pdf the quantity `extension' of the pdf:

$$w_\xi = \text{ext}\left\{ f_\xi(x) \right\} = b - a$$

then the general result valid for uniformly-distributed random variables is:

$$\sigma_\xi^2 = \frac{w_\xi^2}{12}$$

Eq. 11-34

The standard deviation is obviously:

$$\sigma_\xi = \frac{w_\xi}{2\sqrt{3}}$$

Note also that the overall pdf can be expressed as:

$$f_\xi(x) = \frac{1}{w_\xi} \Pi_{w_\xi}(x - \mu_\xi)$$

Eq. 11-35

or also:

$$f_\xi(x) = \frac{1}{2\sqrt{3\sigma_\xi^2}} \Pi_{2\sqrt{3\sigma_\xi^2}}(x - \mu_\xi)$$

11.4.2.4 Example: mean and variance of a Gaussian RV

We now look at a Gaussian distribution:

$$f_\xi(x) = \frac{1}{\sqrt{2\pi\sigma_\xi^2}} e^{-\frac{(x-\mu_\xi)^2}{2\sigma_\xi^2}}$$

Eq. 11-36

This pdf explicitly contains its mean and variance, as parameters.

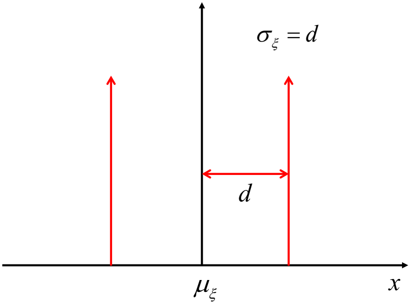
Demonstrating that the mean value is indeed μ_ξ is quite easy and can be done on your own. Hint: write down the integral formula and then change integration variable, from x to $z = x - \mu_\xi$. The you can split up the integral into two integrals, one of which is certainly 0 because its integrand function is odd.

Proving that the standard deviation is σ_ξ^2 is less obvious and requires rather complex integration procedures or the use of a table of integrals.

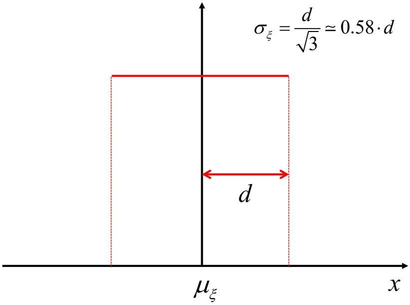
11.4.3 relationship of standard deviation vs. "width" of pdf

As mentioned, the standard deviation roughly relates to how "spread-out" a distribution is, about its mean value. For the examples of pdf's we provide here in formulas and pictorial form the actual relationship.

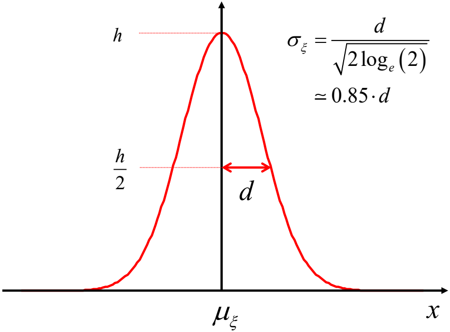
Note that designating the width "d" is rather arbitrary. Other choices are possible, but the ones proposed below are rather intuitive.



$\sigma_\xi = d$
d
μ_ξ
x



$\sigma_\xi = \frac{d}{\sqrt{3}} \simeq 0.58 \cdot d$
d
μ_ξ
x



h
$\frac{h}{2}$
$\sigma_\xi = \frac{d}{\sqrt{2\log_e(2)}}$
$\simeq 0.85 \cdot d$
d
μ_ξ
x

$\boxed{\text{On your own:}}$ verify that the three results reported above are correct.

11.4.3.1 more examples of pdf's (optional)

There are of course infinitely many possible pdf's. Apart from the Gaussian and uniform pdf's already presented, famous ones are the Rayleigh and Rice distributions. These are in fact both obtained from Gaussian RV's. Specifically, given two statistically independent Gaussian RV's, ξ and η, with same variance σ^2, we define a new RV ρ as:

$$\rho = \sqrt{\xi^2 + \eta^2}$$

If ξ and η are both zero-mean, then a Rayleigh distribution is obtained:

$$f_\rho(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} u(x)$$

The mean and variance of a Rayleigh-distributed RV are:

$$\mu_\rho = \sigma\sqrt{\pi/2} \qquad \sigma_\rho^2 = \left(2-\pi/2\right)\sigma^2$$

If ξ and η are non-zero mean, then a *Rice* distribution is obtained (details omitted). Both these distributions are important for the analysis of telecommunication systems, but they also appear in many other contexts, both in engineering and physics.

(end of optional)

11.4.4 (optional) Examples of calculating probabilities of intervals in $\mathbb{R}$

As mentioned, given a RV ξ and its pdf $f_\xi(x)$, all probabilities of the type $P(\{\xi \in A\})$, where $A \subseteq \mathbb{R}$, can be calculated using the fundamental formula:

$$P\left(\{\xi \in A\}\right) = \int_A f_\xi(x)\,dx$$

We will concentrate on the simple case where the subset of $\mathbb{R}$ is an interval $\mathbf{I} = [x_1, x_2]$. We then have:

$$P\left(\{\xi \in \mathbf{I}\}\right) = \int_{x_1}^{x_2} f_\xi(x)\,dx$$

11.4.4.1 (optional) uniform pdf and probability of intervals

If we consider a uniformly distributed RV, then the result is:

$$P\left(\left\{\xi \in \mathbf{I}\right\}\right) = \int_{x_1}^{x_2} f_\xi\left(x\right) dx$$

$$= \frac{1}{w_\xi} \int_{x_1}^{x_2} \Pi_{w_\xi}\left(x - \mu_\xi\right) dx = \frac{w_{\text{int}}}{w_\xi}$$

where $w_{\text{int}} = x_h - x_l$, with:

$$\left[x_l, x_h\right] = \mathbf{I} \cap \operatorname{supp}\left\{\Pi_{w_\xi}\left(x - \mu_\xi\right)\right\} = \left[x_1, x_2\right] \cap \left[\mu_\xi - w_\xi/2, \mu_\xi + w_\xi/2\right]$$

If the above intersection is empty, then $w_{\text{int}} = 0$.

(end of optional)

11.4.4.2 Gaussian pdf and probability of intervals

If we consider a Gaussian-distributed RV, then the result is:

$$P\left(\{\xi \in \mathbf{I}\}\right) = \int_{x_1}^{x_2} f_\xi\left(x\right) dx = \frac{1}{\sqrt{2\pi\sigma_\xi^2}} \int_{x_1}^{x_2} e^{-\frac{\left(x-\mu_\xi\right)^2}{2\sigma_\xi^2}} dx$$

$$= \frac{1}{2}\operatorname{erf}\left(\frac{x_2-\mu_\xi}{\sqrt{2}\,\sigma_\xi}\right) - \frac{1}{2}\operatorname{erf}\left(\frac{x_1-\mu_\xi}{\sqrt{2}\,\sigma_\xi}\right)$$

The result stems from the indefinite integral of the Gaussian pdf formula:

$$\frac{1}{\sqrt{2\pi\sigma_\xi^2}} \int e^{-\frac{\left(x-\mu_\xi\right)^2}{2\sigma_\xi^2}} dx = \frac{1}{2}\operatorname{erf}\left(\frac{x_1-\mu_\xi}{\sqrt{2}\,\sigma_\xi}\right) + C$$

The “erf” function, or “error function”, is a special function. However, it is built into many scientific calculators and it is pre-defined on most of the popular scientific software, such as Matlab.

11.4.4.3 Rayleigh pdf (optional)

Considering the Rayleigh pdf, the result is:

$$P\left(\left\{\xi \in \mathbf{I}\right\}\right) = \int_{x_1}^{x_2} f_{\xi}\left(x\right) dx = \int_{x_1}^{x_2} \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} u\left(x\right) dx$$

$$= \begin{cases} e^{-\frac{x_1^2}{2\sigma^2}} - e^{-\frac{x_2^2}{2\sigma^2}} \quad \text{if } x_1 \geq 0, x_2 \geq 0 \\ 1 - e^{-\frac{x_2^2}{2\sigma^2}} \quad \text{if } x_1 < 0, x_2 \geq 0 \\ 0 \quad \text{if } x_1 \geq 0, x_2 \geq 0 \end{cases}$$

(end of optional)

11.4.4.4 Problem (transfer of expectation)

In this problem we look at the use of the transfer of expectation property of the expectation operator.

We assume that ξ is defined as:

$$\xi = \cos(\varphi)$$

We then assume that φ is a uniformly-distributed random variable in the range $[-\pi/2, \pi/2]$. We want to find the mean value and standard deviation of ξ. This is done as follows.

By definition:

$$\begin{aligned} \mu_\xi &= E\{\xi\} \\ \sigma_\xi^2 &= E\{\xi^2\} - \mu_\xi^2 \end{aligned}$$

But we do not have the pdf of ξ, so we cannot calculate these expectations directly. We can however use Eq. 11-24 and obtain for the mean value:

$$E_\xi\{\xi\} = E_\varphi\{q(\varphi)\} = E_\varphi\{\cos(\varphi)\} = \int_{-\infty}^{\infty} \cos(z) f_\varphi(z)\, dz$$

We now remark that: $f_\varphi(z) = \Pi_\pi(z)/\pi$, so:

$$\mu_\xi = E_\xi\{\xi\} = \tfrac{1}{\pi}\int_{-\pi/2}^{\pi/2} \cos(z)\, dz = \tfrac{1}{\pi}\left[\sin(z)\right]_{-\pi/2}^{\pi/2} = \tfrac{2}{\pi}$$

We now proceed with calculating the variance.

$$\sigma_\xi^2 = E\left\{\xi^2\right\} - \tfrac{4}{\pi^2}$$

So, we only need to compute the mean square value $E\left\{\xi^2\right\}$ Using the same procedure as before, we write:

$$E_\xi\left\{\xi^2\right\} = E_\varphi\left\{\cos^2(\varphi)\right\} = \int_{-\infty}^{\infty}\cos^2(z)\, f_\varphi(z)\, dz =$$

$$= \tfrac{1}{\pi}\int_{-\pi/2}^{\pi/2}\cos^2(z)\, dz = \tfrac{1}{\pi}\int_{-\pi/2}^{\pi/2}\frac{1+\cos(2z)}{2}\, dz =$$

$$= \tfrac{1}{2\pi}\int_{-\pi/2}^{\pi/2} 1(z)dz + \tfrac{1}{2\pi}\int_{-\pi/2}^{\pi/2}\cos(2z)\, dz = \frac{1}{2\pi}\cdot\pi = \frac{1}{2}$$

Note that the last integral, involving $\cos(2z)$, is zero. In fact:

$$\int_{-\pi/2}^{\pi/2}\cos(2z)\, dz = \tfrac{1}{2}\int_{-\pi}^{\pi}\cos(y)\, dy = 0$$

where we have changed variable: $2z = y$, $dz = dy/2$. The integral is zero because it integrates a cosine over its full period 2π.

The final result is then: $\sigma_\xi^2 = \frac{1}{2} - \frac{4}{\pi^2} \approx 0.1$

11.5 Two jointly distributed random variables

11.5.1 The joint pdf

We denote the joint pdf of two random variables as:

$$f_{\xi\eta}(x, y)$$

Similar to the pdf of a single RV, this quantity can be given the interpretation of a "density of probability" versus the possible results of the random experiment, labeled by the numbers x and y, for ξ and η, respectively.

11.5.2 Optional

To show how to get to it, we first introduce the intervals along the x and y axes:

$$\Delta x_i = \left[x_i - \frac{\Delta x}{2}, x_i + \frac{\Delta x}{2} \right]$$

$$\Delta y_j = \left[y_j - \frac{\Delta y}{2}, y_j + \frac{\Delta y}{2} \right]$$

where x_i and y_j represent an arbitrary point (x_i, y_j) over the plane (x, y). We can then write the probability that the result of the experiment ends up in Δx_i for ξ *and* simultaneously ends up in Δy_j for η, as:

$$P\left(\left\{\xi \in \Delta x_i, \eta \text{ any}\right\} \bigcap \left\{\eta \in \Delta y_j, \xi \text{ any}\right\}\right)$$

Eq. 11-37

Note that this probability is of the form $P(A \bigcap B)$, where A and B are the composite events $A = \{\xi \in \Delta x_i, \eta \text{ any}\}$ and $B = \{\eta \in \Delta y_j, \xi \text{ any}\}$. So Eq. 11-37 is a so-called *joint probability* of the events A and B.

To get to a probability density, we formally divide Eq. 11-37 by Δx and Δy and take the limit:

$$f_{\xi\eta}\left(x_i, y_j\right) = \lim_{\Delta x, \Delta y \to 0} \frac{P\left(A \bigcap B\right)}{\Delta x \Delta y}$$

$$= \lim_{\Delta x, \Delta y \to 0} \frac{P\left(\left\{\xi \in \Delta x_i, \eta \text{ any}\right\} \bigcap \left\{\eta \in \Delta y_j, \xi \text{ any}\right\}\right)}{\Delta x \Delta y}$$

Eq. 11-38

This can actually be done at any point over the plane, so that we can drop the indices i and j and simply write $f_{\xi\eta}\left(x, y\right)$ as:

$$f_{\xi\eta}\left(x, y\right) = \lim_{\Delta x, \Delta y \to 0} \frac{P\left(A \bigcap B\right)}{\Delta x \Delta y}$$

$$= \lim_{\Delta x, \Delta y \to 0} \frac{P\left(\left\{\xi \in \mathbf{\Delta} x, \eta \text{ any}\right\} \bigcap \left\{\eta \in \mathbf{\Delta} y, \xi \text{ any}\right\}\right)}{\Delta x \Delta y}$$

where:

$$\mathbf{\Delta}x = \left[x - \frac{\Delta x}{2}, x + \frac{\Delta x}{2} \right]$$

$$\mathbf{\Delta}y = \left[y - \frac{\Delta y}{2}, y + \frac{\Delta y}{2} \right]$$

(end of optional material)

11.5.3 Meaning of the joint pdf

Remembering that in the single-variable case we wrote:

$$f_\xi(x) = \lim_{\Delta x \to 0} \frac{P\left(\left\{ \xi \in \mathbf{\Delta}x \right\}\right)}{\Delta x} = \frac{dP}{dx}$$

$$\text{with:} \quad \mathbf{\Delta}x = \left[x - \frac{\Delta x}{2}, x + \frac{\Delta x}{2} \right]$$

we can then similarly write:

$$f_{\xi\eta}(x,y) = \lim_{\Delta x,\Delta y \to 0} \frac{P\left(\left\{\xi \in \mathbf{\Delta} x\right\} \cap \left\{\eta \in \mathbf{\Delta} y\right\}\right)}{\Delta x \Delta y} = \frac{d^2 P}{dx dy}$$

$$\text{with:} \quad \begin{array}{l} \mathbf{\Delta} x = \left[x - \dfrac{\Delta x}{2}, x + \dfrac{\Delta x}{2} \right] \\ \mathbf{\Delta} y = \left[y - \dfrac{\Delta y}{2}, y + \dfrac{\Delta y}{2} \right] \end{array}$$

For the single-variable case, we concluded that we could therefore think of $f_\xi(x)$ as a linear density of probability over the $\mathbb{R}$ at the point x. In the two-variable case, we can think of $f_{\xi\eta}(x,y)$ as the *surface density of probability* over the plane $\mathbb{R}^2$ at the point (x,y).

Then, if one wants to know what the probability is of one of the random experiment result to end up in a rectangular interval over the (x, y) plane, defined by a generic $\mathbf{\Delta}x$ and $\mathbf{\Delta}y$, such probability can be calculated by integrating over that rectangular interval:

$$P\left(\left\{\xi \in \mathbf{\Delta}x\right\} \cap \left\{\eta \in \mathbf{\Delta}y\right\}\right) = \int_{\mathbf{\Delta}x} \int_{\mathbf{\Delta}y} f_{\xi\eta}(x, y)\, dy dx$$

Intervals however need not be rectangular. The probability of the result of the random experiment ending up within *any arbitrary two-dimensional domain* $\wp$ *of any shape* over the (x, y) plane, is similarly found as:

$$P\left(\left\{(\xi, \eta) \in \wp\right\}\right) = \iint_{\wp} f_{\xi\eta}(x, y)\, dx dy$$

Since the whole plane is the entire space of events, that is $\Omega = \mathbb{R}^2$, integrating over the whole of it must return probability 1:

$$P(\Omega) = P\left(\mathbb{R}^2\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\xi\eta}(x, y)dxdy = 1$$

Note also that probabilities must always be positive. So, without exception, it must be:

$$f_{\xi\eta}(x, y) \geq 0$$

11.5.4 conditional and marginal pdf's

The joint pdf can be written as the product:

$$f_{\xi,\eta}(x,y) = f_{\xi|\eta}(x\,|\,y)\,f_\eta(y) = f_{\eta|\xi}(y\,|\,x)\,f_\xi(x)$$

Eq. 11-39

where $f_{\xi|\eta}(x\,|\,y)$ and $f_{\eta|\xi}(y\,|\,x)$ are the *conditional pdf's*.

This formula bears a striking similarity to the well-known Bayes' formula:

$$P(A \cap B) = P(A\,|\,B)\,P(B) = P(B\,|\,A)\,P(A)$$

Eq. 11-40

In fact, Eq. 11-39 can be proved based on Eq. 11-40. This is done below, but it is left as optional material.

Keep in mind though that conditional pdf's like $f_{\xi|\eta}(x\,|\,y)$ and $f_{\eta|\xi}(y\,|\,x)$ are pdf's to all effects. In particular, they integrate to 1, are always positive and provide the probability of the first random variable in the subscript to show up in a given interval, *given that the other RV in the subcript has acquired a certain value*.

For instance:

$$\int_{\mathrm{I}} f_{\xi|\eta}(x \mid y)\, dx = \mathrm{P}\{\xi \in \mathrm{I} \mid \text{given } \eta = y\}$$

is the probability that ξ ends up in I, given that the result for η has been a certain number y.

Optional **–** conditional pdf's from Bayes' formula

We start out from:

$$P(A \cap B) = P(A \mid B)\, P(B)$$

which we use to rewrite Eq. 11-38 as follows:

$$f_{\xi\eta}\left(x_i, y_j\right) = \lim_{\Delta x, \Delta y \to 0} \frac{P\left(A \cap B\right)}{\Delta x \Delta y} = \lim_{\Delta x, \Delta y \to 0} \frac{P\left(A \mid B\right) P\left(B\right)}{\Delta x \Delta y}$$

Eq. 11-41

The quantity $P\left(A \mid B\right)$ clearly is:

$$\begin{aligned} P\left(A \mid B\right) &= P\left(\left\{\xi \in \Delta x_i,\ \eta \text{ any} \mid \eta \in \Delta y_j\ ,\ \xi \text{ any}\right\}\right) \\ &= P\left(\left\{\xi \in \Delta x_i \mid \eta \in \Delta y_j\right\}\right) \end{aligned}$$

Eq. 11-42

where the first equality is found by definition of $A \mid B$. The second equality is found by removing "ξ any" because conditioning on "any" result for ξ means no conditioning, and by removing "η any" because the *conditioning* on $\eta \in \Delta y_j$ makes specifying "η any" irrelevant.

Substituting Eq. 11-42 into Eq. 11-41 we then get:

$$f_{\xi\eta}\left(x_i, y_j\right) = \lim_{\Delta x, \Delta y \to 0} \frac{P\left(\left\{\xi \in \Delta x_i \mid \eta \in \Delta y_j\right\}\right)}{\Delta x} \frac{P(B)}{\Delta y}$$

Eq. 11-43

We can then take the limit for $\Delta y \to 0$, on those quantities that depend on Δy, which yields:

$$\lim_{\Delta y \to 0} P\left(\left\{\xi \in \Delta x_i \,|\, \eta \in \Delta y_j\right\}\right) \frac{P(B)}{\Delta y} = P\left(\left\{\xi \in \Delta x_i \,|\, \eta = y_j\right\}\right) f_\eta\left(y_j\right)$$

Eq. 11-44

The above result is found by remarking that:

$$\lim_{\Delta y \to 0} P\left(\left\{\xi \in \Delta x_i \,|\, \eta \in \Delta y_j\right\}\right) = P\left(\left\{\xi \in \Delta x_i \,|\, \eta = y_j\right\}\right)$$

and that:

$$\begin{aligned}\lim_{\Delta y \to 0} \frac{P(B)}{\Delta y} &= \lim_{\Delta y \to 0} \frac{P\left(\left\{\eta \in \Delta y_j, \xi \text{ any}\right\}\right)}{\Delta y} \\ &= \lim_{\Delta y \to 0} \frac{P\left(\left\{\eta \in \Delta y_j\right\}\right)}{\Delta y} = f_\eta\left(y_j\right)\end{aligned}$$

by definition of pdf.

Substituting Eq. 11-44 into Eq. 11-43, we then get:

$$f_{\xi\eta}\left(x_i, y_j\right) = f_\eta\left(y_j\right) \cdot \lim_{\Delta x \to 0} \frac{P\left(\left\{\xi \in \Delta x_i \,|\, \eta = y_j\right\}\right)}{\Delta x}$$

Comparing it with Eq. 11-39 we see that it must be:

$$f_{\xi|\eta}\left(x_i \,|\, y_j\right) = \lim_{\Delta x \to 0} \frac{P\left(\left\{\xi \in \Delta x_i \,|\, \eta = y_j\right\}\right)}{\Delta x}$$

This formula represents a definition of the conditional pdf of ξ given η which makes its meaning explicit. It shows that it is the (one-dimensional) pdf of the RV ξ, *given that the joint random experiment has resulted in* $\eta = y_j$. We finally remark that the coordinates x_i and y_j are arbitrary and can thus be replaced by simply x

and y.

Like all pdf's, conditional pdf's too are always positive or at least zero, and must integrate to 1:

$$\int_{-\infty}^{\infty} f_{\xi|\eta}(x \mid y)dx = 1$$

$$\int_{-\infty}^{\infty} f_{\eta|\xi}(y \mid x)dy = 1$$

Note that, similar to to the intepretation of $P(A \mid B)$ as $P_B(A)$, that is $P(A \mid B)$ being the probability of A in a reduced space of events that consists of just B, the quantity $f_{\xi|\eta}(x \mid y)$ can be seen as the probability density of ξ in a reduced space of events consisting of $\eta = y$, that is no longer the plane $\mathbb{R}^2$ but just a single line with a fixed coordinate y and a free coordinate x.

End of optional material.

11.5.4.1 Statistical Independence (important)

If ξ and η are statistically independent, then:

$$f_{\xi|\eta}(x|y) = f_\xi(x)$$
$$f_{\eta|\xi}(y|x) = f_\eta(y)$$

Eq. 11-45

(*Optional:* the above formulas are found because, if A and B are statistically independent, then $P(A|B) = P(A)$ in Eq. 11-41. From there, it is easy to derive Eq. 11-45. *End of optional.*)

Using Eq. 11-45 in Eq. 11-39, it is immediately found that statistical independence also implies the following key result:

$$f_{\xi\eta}(x,y) = f_{\xi}(x) f_{\eta}(y)$$

11.5.4.2 marginal pdf's

The individual pdf's of the two RV described by a joint pdf $f_{\xi\eta}(x,y)$, that is $f_{\xi}(x)$ and $f_{\eta}(y)$, are called *marginal pdf's*. Quite interestingly, they can be found from the joint pdf $f_{\xi\eta}(x,y)$ by "integrating out" the other RV:

$$f_{\eta}(y) = \int_{-\infty}^{\infty} f_{\xi\eta}(x,y)dx$$

$$f_{\xi}(x) = \int_{-\infty}^{\infty} f_{\xi\eta}(x,y)dy$$

This shows that all the statistical information regarding each single RV *alone* is available within the joint pdf which, in addition, also contains the statistical dependence between the two.

11.5.4.3 Optional approximating the joint pdf

The joint pdf $f_{\xi\eta}(x,y)$ can be *approximated* by resorting to a "3D histogram", where the "bins" are prisms with a square base (also called "right parallelepipeds[2] with a square base"), built as follows. First, we estimate $f_{\xi\eta}(x,y)$ at certain specific values of x and y, say x_i and y_j, which we choose equally spaced along the two real axes, with spacing $\Delta x = \Delta y = \Delta$. To do so, we define

$$\Delta x_i = \left[x_i - \frac{\Delta}{2}, x_i + \frac{\Delta}{2} \right]$$

$$\Delta y_j = \left[y_j - \frac{\Delta}{2}, y_j + \frac{\Delta}{2} \right]$$

Then:

[2] Pronunciation: /ˌparəlɛləˈpʌɪpɛd/

$$f_{\xi\eta}\left(x_i, y_j\right) \approx \frac{n_{\{\xi\in\Delta x_i\}\cap\{\eta\in\Delta y_j\}}}{N\Delta^2}$$

where $n_{\{\xi\in\Delta x_i\}\cap\{\eta\in\Delta y_j\}}$ has the meaning of "the number of times the random experiment (ξ, η) generated a result (x, y) which ended up in the square interval identified by Δx_i and Δy_j" and N is the total number of times the random experiment was run.

The histogram-like approximation of the pdf can then be built as follows:

$$f_{\xi\eta}(x, y) \approx \tilde{f}_{\xi\eta}(x, y) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \frac{n_{\{\xi\in\Delta x_i\}\cap\{\eta\in\Delta y_j\}}}{N\Delta^2} \Pi_\Delta(x - x_i) \Pi_\Delta\left(y - y_j\right)$$

This approximation of $f_{\xi\eta}(x, y)$ has the nice feature that it always integrates to 1, even though it is just a histogram approximation. This is true for any number of trials N and for any size of Δ. You can easily prove this on your own.

To actually get $f_{\xi\eta}(x,y)$ one should ideally repeat the experiment an infinity of times *and* reduce Δ to zero. **(end of optional material)**

11.5.4.4 joint discrete RV's

When the random experiment produces a countable set of *discrete results* in $\mathbb{R}^2$, that is $\Omega = \{(x_k, y_k)\}_{k=1}^{K}$, where K could also be infinity (but countable), the resulting pdf has the form:

$$f_{\xi\eta}(x,y) = \sum_{k=1}^{K} P\left(\{(\xi,\eta)=(x_k,y_k)\}\right)\delta(x-x_k)\delta(y-y_k)$$

where $P\left(\{(\xi,\eta)=(x_k,y_k)\}\right)$ is the probability of the result being (x_k,y_k). Such pdf complies with all the theoretical requirements of a conventional joint pdf.

As an example, the joint pdf generated by rolling two dice independently is:

$$f_{\xi\eta}(x,y)=\sum_{m=1}^{6}\sum_{n=1}^{6}\frac{1}{36}\delta(x-m)\delta(y-n)$$

Clearly this pdf can be factorized into a factor that depends only on x and one that depends only on y, proving that indeed ξ and η are statistically independent:

$$f_{\xi\eta}(x,y)=\left[\sum_{n=1}^{6}\frac{1}{6}\delta(x-m)\right]\cdot\left[\frac{1}{6}\sum_{m=1}^{6}\delta(y-n)\right]=f_{\xi}(x)\cdot f_{\eta}(y)$$

This pdf can be plotted in 3D or it can be plotted in 2D using a "dot" to indicate where a delta is located in the (x,y) plane. For the specific case of independently rolling two dice, the pdf is:

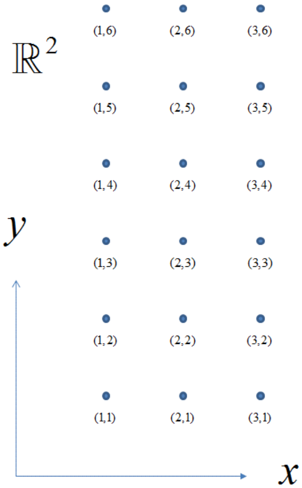


Fig. 11-3 Each dot represents a "2D delta". In this case, each delta has a coefficient $1/36$.

11.5.4.5 example of statistical dependence for continuous random variables

One example of statistical dependence is the following. We are looking at two jointly-distributed RV's ξ and η, where ξ is uniformly distributed between -1/2 and 1/2, and η is defined as:

$$\eta = \xi + \rho$$

where ρ is uniformly distributed between -1/2 and 1/2 and is independent of ξ.

To find the joint pdf of ξ and η, we are going to use the formula $f_{\xi\eta}(x, y) = f_{\eta|\xi}(y \mid x) f_{\xi}(x)$, as follows.

First, $f_{\xi}(x) = \Pi(x)$ by assumption. Then, we consider $f_{\eta|\xi}(y \mid x)$. This is the distribution of η given that ξ has taken on the value $\xi = x$. To formally stress the fact that x represents a sure and specific result, we set $\xi = x = x_0$ and look for $f_{\eta|\xi}(y \mid x_0)$. Under this assumption, we have:

$$\eta = \rho + \xi\Big|_{=x_0} = \rho + x_0$$

and therefore the distribution of η is the same as that of ρ, but with an additional rigid translation of x_0. Remembering that ρ is uniform between -½ and ½, and therefore its pdf is $f_\rho(z) = \Pi(z)$, by shifting all possible results by x_0 we get:

$$f_{\eta|\xi}(y \mid x_0) = \Pi(y - x_0)$$

Putting together the two results, we finally find:

$$f_{\xi\eta}(x,y) = f_{\eta|\xi}(y \mid x) f_{\xi}(x) = \Pi_1(x)\Pi_1(y-x)$$

It is immediately evident that $f_{\xi\eta}(x,y)$ cannot be factorized into a factor that depends only on x and one that depends only on y, showing that indeed ξ and η are statistically *dependent*.

We then want to check that:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\xi\eta}(x,y)dxdy = 1$$

Indeed:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Pi_1(x)\Pi_1(y-x)dxdy = \int_{-\infty}^{\infty} \Pi_1(y-x)\int_{-\infty}^{\infty} \Pi_1(x)dxdy = \int_{-\infty}^{\infty} \Pi_1(y-x)dy = 1$$

We also want to find the *marginals*. We start with:

$$f_\eta(y) = \int_{-\infty}^{\infty} f_{\xi\eta}(x,y)dx = \int_{-\infty}^{\infty} \Pi(x)\Pi(y-x)dx$$

Incidentally, we notice that the above integral can be written as a convolution product, of which we know the result :

$$\Pi(y) * \Pi(y) = \Lambda(y)$$

So we have:

$$f_\eta(y) = \Lambda(y)$$

As for the other marginal:

$$f_\xi(x) = \int_{-\infty}^{\infty} f_{\xi\eta}(x,y)dy = \int_{-\infty}^{\infty} \Pi(x)\Pi(y-x)dy$$

$$= \Pi(x)\int_{-\infty}^{\infty} \Pi(y-x)dy = \Pi(x)$$

Finally, we want to find the conditional pdf's: $f_{\xi|\eta}(x\,|\,y) \quad , \quad f_{\eta|\xi}(y\,|\,x)$

These are immediately found using the general formula:

$$f_{\xi\eta}(x,y) = f_{\eta|\xi}(y\,|\,x)\,f_\xi(x) = f_{\xi|\eta}(x\,|\,y)\,f_\eta(y)$$

We have:

$$f_{\xi|\eta}(x\,|\,y) = \frac{f_{\xi\eta}(x,y)}{f_\eta(y)} = \frac{\Pi_1(x)\Pi_1(y-x)}{\Lambda_1(y)}$$

$$f_{\eta|\xi}(y\,|\,x)=\frac{f_{\xi\eta}(x,y)}{f_{\xi}(x)}=\Pi_1(y-x)$$

These are again pdf's of single RV and they have to integrate to 1. In fact:

$$\int_{-\infty}^{\infty} f_{\xi|\eta}(x\,|\,y)dx=\int_{-\infty}^{\infty}\frac{\Pi_1(x)\Pi_1(y-x)}{\Lambda_1(y)}dx=$$

$$=\frac{1}{\Lambda_1(y)}\int_{-\infty}^{\infty}\Pi_1(x)\Pi_1(y-x)dx=\frac{\Lambda_1(y)}{\Lambda_1(y)}=1$$

$$\int_{-\infty}^{\infty} f_{\eta|\xi}(y\,|\,x)dy=\int_{-\infty}^{\infty}\Pi_1(y-x)dy=1$$

$\boxed{\text{On your own:}}$ draw the joint, marginal and conditional pdf's. Try to make sense of the results.

11.5.4.6 $\boxed{\text{On your own:}}$ example of statistical dependence for discrete random variables

We also provide one simple example of statistical dependence between two discrete RV ξ and η. The random experiment produces a result which is the pair of numbers (x, y) and is defined as follows:

- *the first die is rolled and the result is* x
- *the second die is rolled and the result is* z
    - *if both* x *and* z *are even, then* $y = z$
    - *if both* x *and* z *are odd, then* $y = z$
    - *if neither of the above, the second die is rolled again, till one of the two above conditions is met*

It is easy to see that the set of possible results are the points on the plane shown as dots:

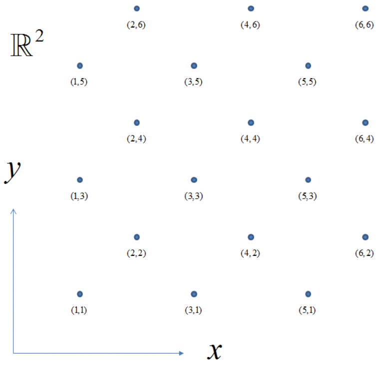


Fig. 11-4: Each dot represents a delta. In this case, each delta has a coefficient $1/18$ **.**

We can also view this as a plot of the pdf, assuming that each dot is a delta. All 18 elementary events are equiprobable, so the deltas are each multiplied times a factor $1/18$.

This is similar to Fig. 11-3, except the first and the second number in the pairs are both even or both odd. The other combinations do not show up: there are no results where one number is odd and the other number is even.

The corresponding analytical pdf formula is:

$$f_{\xi\eta}(x,y) = \sum_{m=1,3,5} \sum_{n=1,3,5} \frac{1}{18} \delta(x-m)\delta(y-n) +$$
$$+ \sum_{m=2,4,6} \sum_{n=2,4,6} \frac{1}{18} \delta(x-m)\delta(y-n)$$

It is easy to see that this pdf cannot be factorized into a factor that depends only on x and one that depends only on y. So ξ and η are not independent.

The two marginals can be found by integration:

$$f_\xi(x) = \int_{-\infty}^{\infty} f_{\xi\eta}(x,y)\,dy = \sum_{m=1}^{6} \frac{1}{6}\delta(x-m)$$

$$f_\eta(y) = \int_{-\infty}^{\infty} f_{\xi\eta}(x,y)\,dx = \sum_{n=1}^{6} \frac{1}{6}\delta(y-n)$$

Prove these results $\boxed{\text{on your own}}$:

One of the conditional pdf's is:

$$f_{\xi|\eta}(x\,|\,y) = \begin{cases} \text{if } y \text{ is 1,3,5:} \quad \displaystyle\sum_{n=1,3,5} \frac{1}{3}\delta(x-n) \\ \text{if } y \text{ is 2,4,6:} \quad \displaystyle\sum_{n=2,4,6} \frac{1}{3}\delta(x-n) \end{cases}$$

and it clearly always integrates to 1. The other is identical, it is enough to exchange ξ with η, and x with y. (end of "on your own")

11.5.5 Higher order pdf's

Joint pdf's can be written for more than two RV's. It can actually be for three:

$$f_{\xi\eta\rho}(x,y,z)$$

Here too the interpretation of this quantity is that of a "density of probability" versus the possible results of the random experiment, labeled by the numbers x, y and z. Namely:

$$f_{\xi\eta\rho}(x,y,z) = \frac{d^3 P}{dxdydz}$$

If one wants to know what the probability is of the random experiment result to end up in any arbitrary three-dimensional domain $\wp$ in the (x, y, z) plane, then such probability is found as:

$$P\left(\left\{(\xi, \eta, \rho) \in \wp\right\}\right) = \iiint_{\wp} f_{\xi\eta\rho}(x, y, z)\, dxdydz$$

Of course, here too:

$$f_{\xi\eta\rho}(x, y, z) \geq 0$$

$$P(S) = P\left(\mathbb{R}^3\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\xi\eta\rho}(x, y, z)\, dxdydz = 1$$

We will not go any deeper into this topic. We only point out that if the three RV's are statistically independent, then:

$$f_{\xi\eta\rho}(x,y,z) = f_{\xi}(x)\, f_{\eta}(y)\, f_{\rho}(z)$$

It is of course also possible to consider joint pdf's of more than three RV's, and in fact, of *any number* of RV's. In this case it is more convenient to use a vector notation. Assuming N RV's, that is: $\xi_1, \xi_2, \ldots, \xi_N$, then their joint pdf can be formally written as:

$$f_{\boldsymbol{\xi}}(\mathbf{x})$$

where:

$$\boldsymbol{\xi} = \left[\xi_1, \xi_2, \ldots, \xi_N\right] \qquad \mathbf{x} = \left[x_1, x_2, \ldots, x_N\right]$$

The key properties are still:

$$f_{\boldsymbol{\xi}}(\mathbf{x}) > 1 \quad ,$$

$$P\left(\left\{\boldsymbol{\xi} \in \wp\right\}\right) = \int_{\wp} f_{\boldsymbol{\xi}}(\mathbf{x})\, d\mathbf{x} \quad , \quad \wp \subseteq \mathbb{R}^N$$

and

$$P(\Omega) = P\left(\left\{\boldsymbol{\xi} \in \mathbb{R}^N\right\}\right) = \int_{\mathbb{R}^N} f_{\boldsymbol{\xi}}(\mathbf{x})d\mathbf{x} = 1$$

11.5.6 The expectation operator for two RV

So far, we have looked at the expectation operator over one RV. We now extend the operator to two RVs.

By definition, the *joint expectation operator* of a generic function of two RVs is:

$$E_{\xi\eta}\{g(\xi,\eta)\} \triangleq \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y) f_{\xi\eta}(x,y)dxdy$$

Eq. 11-46

This operator retains all the usual properties that it has over one random variable, that is: linearity, expectation of an independent quantity and transfer of expectation.

What is new in this context is the fact that the two RVs ξ and η may have different relationships, from total dependence to partial dependence to total independence. In some cases, the joint operator can then be simplified.

We start this discussion with an important general property of joint expectation operators.

11.5.6.1 Conditional expectations

We recall the following key result regarding joint RVs:

$$f_{\xi\eta}(x,y) = f_{\xi|\eta}(x\,|\,y)f_{\eta}(y) = f_{\eta|\xi}(y\,|\,x)f_{\xi}(x)$$

Eq. 11-47

where $f_{\xi|\eta}(x\,|\,y)$ and $f_{\eta|\xi}(y\,|\,x)$ are called *conditional pdfs.*

A conditional pdf is in fact a single RV pdf, obtained under the assumption that the random experiment has returned a certain result for the other RV. For instance, $f_{\xi|\eta}(x|y)$ is the pdf of ξ alone, under the assumption that the random experiment has given, as a result, $\eta = y$.

Based on $f_{\xi|\eta}(x|y)$, we can define a *conditional expectation operator*:

$$E_{\xi|\eta}\{g(\xi, y)\} \triangleq \int_{-\infty}^{+\infty} g(x, y) f_{\xi|\eta}(x|y) dx$$

Eq. 11-48

Note the key fact that $E_{\xi|\eta}\{g(\xi, y)\}$ is a function of y, which is not a RV but rather is a parameter, indicating the value which is assumed for η. In fact, from a statistical viewpoint, the expectation Eq. 11-48 is a single RV expectation over ξ, such as Eq. 11-22, although conditioned on a given result for η.

Likewise, we can define:

$$E_{\eta|\xi}\{g(x,\eta)\} \triangleq \int_{-\infty}^{+\infty} g(x,y) f_{\eta|\xi}(y\,|\,x)dy$$

Based on these definitions, it can be shown that we can re-write the joint expectation operator as:

$$E_{\xi\eta}\{g(\xi,\eta)\} = E_{\eta}\left\{E_{\xi|\eta}\{g(\xi,\eta)\}\right\} = E_{\xi}\left\{E_{\eta|\xi}\{g(\xi,\eta)\}\right\}$$

Eq. 11-49

This important formula can be proved as follows. We first recognize that:

$$E_{\eta}\left\{E_{\xi|\eta}\{g(\xi,\eta)\}\right\} = E_{\eta}\left\{q\left(\eta\right)\right\}$$

having defined:

$$q(\eta) = E_{\xi|\eta}\{g(\xi,\eta)\}$$

Note that indeed the right hand side depends only on η from the "outside", being ξ an internal "variable", the one the conditional average operates on.

We then have:

$$E_\eta\{q(\eta)\} = \int_{-\infty}^{+\infty} f_\eta(y)q(y)dy = \int_{-\infty}^{+\infty} f_\eta(y)E_{\xi|\eta}\{g(\xi,y)\}dy =$$

$$= \int_{-\infty}^{+\infty} f_\eta(y)\int_{-\infty}^{+\infty} g(x,y)f_{\xi|\eta}(x\,|\,y)dxdy =$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f_{\xi|\eta}(x\,|\,y)f_\eta(y)dxdy =$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f_{\xi\eta}(x,y)dxdy = E_{\xi\eta}\{g(\xi,\eta)\}$$

which indeed proves:

$$E_{\xi\eta}\{g(\xi,\eta)\} = E_{\eta}\left\{E_{\xi|\eta}\{g(\xi,\eta)\}\right\}.$$

The key point in the proof was to perform the substitution:

$$f_{\xi\eta}(x,y) = f_{\xi|\eta}(x \mid y) f_{\eta}(y)$$

Similarly, it is easy to prove that: $E_{\xi\eta}\{g(\xi,\eta)\} = E_{\xi}\left\{E_{\eta|\xi}\{g(\xi,\eta)\}\right\}$. The procedure is identical, we only have to substitute

$$f_{\xi\eta}(x,y) = f_{\eta|\xi}(y \mid x) f_{\xi}(x)$$

when needed.

11.5.6.2 statistical independence

If ξ and η are statistically independent, then Eq. 11-47 becomes simply:

$$f_{\xi\eta}(x,y) = f_{\xi}(x) f_{\eta}(y)$$

As a result, we get:

$$E_{\xi\eta}\{g(\xi,\eta)\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y) f_{\xi\eta}(x,y)dxdy$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y) f_{\xi}(x) f_{\eta}(y)dxdy =$$

$$= \int_{-\infty}^{+\infty} f_{\eta}(y) \int_{-\infty}^{+\infty} g(x,y) f_{\xi}(x)dxdy =$$

$$\int_{-\infty}^{+\infty} f_{\eta}(y) E_{\xi}\{g(\xi,y)\}dy = E_{\eta}\left\{E_{\xi}\{g(\xi,\eta)\}\right\}$$

It is easy to show, by exchanging the order of the integration in the above formula, that we also have:

$$E_{\xi\eta}\{g(\xi,\eta)\} = E_{\xi}\left\{E_{\eta}\{g(\xi,\eta)\}\right\}$$

In conclusion:

$$E_{\xi\eta}\{g(\xi,\eta)\} = E_{\eta}\left\{E_{\xi}\{g(\xi,\eta)\}\right\} = E_{\xi}\left\{E_{\eta}\{g(\xi,\eta)\}\right\}$$

Eq. 11-50

This is an important result which is very useful in practical calculations.

If we can further make the following assumption on the function that needs to be averaged:

$$g(x,y) = g_1(x)g_2(y)$$

that is, if we find that the function $g(x,y)$ can be factorized into a function g_1 that depends only on x and a function g_2 that depends only on y, then we get:

$$E_{\xi\eta}\{g(\xi,\eta)\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y) f_{\xi\eta}(x,y)\,dxdy$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g_1(x) g_2(y) f_\xi(x) f_\eta(y)\,dxdy =$$

$$= \int_{-\infty}^{+\infty} f_\eta(y) g_2(y)\,dy \int_{-\infty}^{+\infty} g_1(x) f_\xi(x)\,dx =$$

$$= E_\xi\left\{g_1(\xi)\right\} E_\eta\left\{g_2(\eta)\right\}$$

In summary:

if ξ and η are statistically independent and $g(x,y) = g_1(x) g_2(y)$, then:

$$E_{\xi\eta}\{g(\xi,\eta)\} = E_{\xi}\left\{g_1(\xi)\right\} E_{\eta}\left\{g_2(\eta)\right\}$$

Eq. 11-51

11.5.6.3 joint moments

The *joint non-central moment of order* $m, n,$ of two random variables ξ and η is defined as follows:

$$E_{\xi\eta}\{\xi^m\eta^n\} \triangleq \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^m y^n f_{\xi\eta}(x,y)dxdy$$

Eq. 11-52

If ξ and η are statistically independent, then, according to Eq. 11-51, we have:

$$E_{\xi\eta}\{\xi^m\eta^n\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^m y^n f_\xi(x) f_\eta(y) dxdy =$$

$$= \int_{-\infty}^{+\infty} x^m f_\xi(x) dx \int_{-\infty}^{+\infty} y^n f_\eta(y) dy = E\{\xi^m\} E\{\eta^n\}$$

Eq. 11-53

As an example, the *joint non-central moment of order 1,1* is:

$$E_{\xi\eta}\{\xi\eta\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x\, y\, f_{\xi\eta}(x, y) dxdy$$

Eq. 11-54

In case of *statistical independence of* ξ *and* η*,* owing to Eq. 11-51, we simply get:

$$E_{\xi\eta}\{\xi\eta\} = E_\xi\{\xi\} E_\eta\{\eta\} = \mu_\xi \mu_\eta$$

Eq. 11-55

The *joint central moment of order* $m,n,$ of two random variables ξ and η is defined as follows:

$$E_{\xi\eta}\left\{\left(\xi-\mu_{\xi}\right)^{m}\left(\eta-\mu_{\eta}\right)^{n}\right\}=$$
$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left(x-\mu_{\xi}\right)^{m}\left(y-\mu_{\eta}\right)^{n} f_{\xi\eta}(x,y)dxdy$$

Eq. 11-56

If ξ and η are statistically independent, then:

$$E_{\xi\eta}\left\{\left(\xi-\mu_\xi\right)^m\left(\eta-\mu_\eta\right)^n\right\}=$$

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left(x-\mu_\xi\right)^m\left(y-\mu_\eta\right)^n f_\xi(x)f_\eta(y)dxdy=$$

$$=\int_{-\infty}^{+\infty}\left(x-\mu_\xi\right)^m f_\xi(x)dx\int_{-\infty}^{+\infty}\left(y-\mu_\eta\right)^n f_\eta(y)dy=$$

$$=E_\xi\left\{\left(\xi-\mu_\xi\right)^m\right\}\cdot E_\eta\left\{\left(\eta-\mu_\eta\right)^n\right\}$$

Eq. 11-57

Note that we could have directly found the above result by using Eq. 11-51.

The joint central moment of order (1,1) is important and is called *covariance*. It is sometimes identified with the symbol $\sigma_{\xi\eta}$ and its definition is:

$$\sigma_{\xi\eta} \triangleq E_{\xi\eta}\left\{\left(\xi-\mu_{\xi}\right)\left(\eta-\mu_{\eta}\right)\right\}$$

Eq. 11-58

An alternative definition for the covariance is easily found by first executing the products inside and then using the linearity of the expectation operator:

$$\begin{aligned}
\sigma_{\xi\eta} &= E_{\xi\eta}\left\{\xi\eta-\mu_{\xi}\eta-\mu_{\eta}\xi+\mu_{\xi}\mu_{\eta}\right\} \\
&= E_{\xi\eta}\left\{\xi\eta\right\}-\mu_{\xi}E_{\eta}\left\{\eta\right\}-\mu_{\eta}E_{\xi}\left\{\xi\right\}+\mu_{\xi}\mu_{\eta} \\
&= E_{\xi\eta}\left\{\xi\eta\right\}-2\mu_{\xi}\mu_{\eta}+\mu_{\xi}\mu_{\eta} \\
&= E_{\xi\eta}\left\{\xi\eta\right\}-\mu_{\xi}\mu_{\eta}
\end{aligned}$$

Eq. 11-59

In case the RVs are statistically independent, using Eq. 11-51, we get:

$$
\begin{aligned}
\sigma_{\xi\eta} &= E_{\xi\eta}\{\xi\eta\} - \mu_\xi \mu_\eta \\
&= E_\xi\{\xi\} E_\eta\{\eta\} - \mu_\xi \mu_\eta \\
&= \mu_\xi \mu_\eta - \mu_\xi \mu_\eta \\
&= 0
\end{aligned}
$$

This proves the following significant result:

given any two statistically independent RVs, their covariance is always equal to zero.

Note however that the converse is not, in general true. That is:

$$
\left\{ \begin{matrix} \text{statistical} \\ \text{independence} \end{matrix} \right\} \Rightarrow \left\{ \sigma_{\xi\eta} = 0 \right\}
$$

but

$$
\left\{ \sigma_{\xi\eta} = 0 \right\} \not\Rightarrow \left\{ \begin{matrix} \text{statistical} \\ \text{independence} \end{matrix} \right\}
$$

11.5.7 The correlation coefficient

A very important quantity derived from joint moments is the so-called *correlation coefficient*:

$$\rho_{\xi\eta} \triangleq \frac{\sigma_{\xi\eta}}{\sigma_\xi \sigma_\eta} = \frac{E_{\xi\eta}\left\{(\xi-\mu_\xi)(\eta-\mu_\eta)\right\}}{\sigma_\xi \sigma_\eta}$$

Eq. 11-60

The following alternative definition is of course equivalent:

$$\rho_{\xi\eta} \triangleq \frac{E_{\xi\eta}\{\xi\eta\}-\mu_\xi \mu_\eta}{\sigma_\xi \sigma_\eta}$$

Eq. 11-61

The correlation coefficient inherits some of the properties of the covariance. In case the RVs are *statistically independent* $\sigma_{\xi\eta} = 0$. Then of course the same is true for $\rho_{\xi\eta}$:

$$\rho_{\xi\eta} = \frac{\sigma_{\xi\eta}}{\sigma_\xi \sigma_\eta} = 0$$

Eq. 11-62

Of course, as for the covariance, the converse is not, in general true. That is:

$$\left\{ \begin{matrix} \text{statistical} \\ \text{independence} \end{matrix} \right\} \Rightarrow \left\{ \rho_{\xi\eta} = 0 \right\}$$

but

$$\left\{ \rho_{\xi\eta} = 0 \right\} \not\Rightarrow \left\{ \begin{matrix} \text{statistical} \\ \text{independence} \end{matrix} \right\}$$

In other words, even though the correlation coefficient between random variables is zero, this does not necessarily mean that they are statistically independent. One can only say that they are *uncorrelated*.

On your own ***Example***

Let's consider for instance a pair ξ and η of RV's. Let's assume that ξ is zero-mean and Gaussian. Let's define η as:

$$\eta = |\xi|$$

It is obvious that ξ and η are statistically dependent. In fact, η is a deterministic function of ξ. However, their correlation coefficient is $\rho_{\xi\eta} = 0$.

On your own: prove that in fact $\rho_{\xi\eta} = 0$, by direct calculation.

Another general result regarding the correlation coefficient is the following:

the absolute value of the correlation coefficient is always less than or at most equal to 1:

$$\left|\rho_{\xi\eta}\right| \leq 1$$

The value 1 is reached in special cases, when the two random variables are deeply connected. In particular, the following result holds:

the correlation coefficient of two RVs ξ *and* η *is* $\rho_{\xi\eta} = \pm 1$ *when the two RVs are related through a relationship of the type:*

$$\eta = a\xi + b$$

The proof of this result is straightforward. We assume we know μ_ξ and σ_ξ^2. The correlation coefficient is as shown in Eq. 11-62, so it requires σ_ξ, σ_η and $\sigma_{\xi\eta}$. Of course, $\sigma_\xi = \sqrt{\sigma_\xi^2}$. Then, to calculate $\sigma_\eta^2 = E_\eta\left\{\left(\eta - \mu_\eta\right)^2\right\}$ we need also μ_η:

$$\mu_\eta = E_\eta\{\eta\} = E_\xi\{a\xi + b\} = a\mu_\xi + b$$

so that:

$$\sigma_\eta^2 = E_\eta\left\{\left(\eta - \mu_\eta\right)^2\right\} = E_\xi\left\{\left(a\xi + b - a\mu_\xi - b\right)^2\right\} =$$

$$= a^2 E_\xi\left\{\left(\xi - \mu_\xi\right)^2\right\} = a^2\sigma_\xi^2$$

and therefore:

$$\sigma_\eta = \sqrt{\sigma_\eta^2} = \sqrt{a^2\sigma_\xi^2} = |a|\sigma_\xi$$

Finally:

$$\sigma_{\xi\eta} = E_\eta\left\{\left(\xi-\mu_\xi\right)\left(\eta-\mu_\eta\right)\right\} = E_\xi\left\{\left(\xi-\mu_\xi\right)\left(a\xi+b-a\mu_\xi-b\right)\right\}$$
$$= E_\xi\left\{\left(\xi-\mu_\xi\right)\left(a\xi-a\mu_\xi\right)\right\} = aE_\xi\left\{\left(\xi-\mu_\xi\right)^2\right\} = a\cdot\sigma_\xi^2$$

Plugging all these result into the correlation coefficient formula, we get:

$$\rho_{\xi\eta} = \frac{\sigma_{\xi\eta}}{\sigma_\xi\sigma_\eta} = \frac{a\cdot\sigma_\xi^2}{\sigma_\xi\cdot|a|\cdot\sigma_\xi} = \frac{a\cdot\sigma_\xi^2}{|a|\cdot\sigma_\xi^2} = \frac{a}{|a|} = \mathrm{sign}\left(a\right)$$

Obviously, the final result is $\rho_{\xi\eta} = \pm 1$, depending on the sign of a, as we wanted to prove.

The correlation coefficient is, in general, a quite imperfect indicator of statistical dependence. However, since in many practical systems *noise* is Gaussian, then in that context it assumes a greater significance thanks to the bi-implication shown in Eq. 11-65. In the following we introduce the joint Gaussian pdf and then we will show that in that case zero correlation also means statistical independence (but only in that case).

11.5.8 Jointly Gaussian RVs

Two RVs ξ_1 and ξ_2 are jointly Gaussian if their joint pdf has the form:

$$f_{\xi_1\xi_2}(x_1,x_2)=$$

$$=\frac{1}{2\pi\cdot\sqrt{\sigma_1^2\sigma_2^2-\sigma_{12}\sigma_{21}}}\exp\left\{-\frac{1}{2}\cdot\left[(x_1-\mu_1)\quad(x_2-\mu_2)\right]\begin{bmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{21} & \sigma_2^2\end{bmatrix}^{-1}\begin{bmatrix}(x_1-\mu_1)\\(x_2-\mu_2)\end{bmatrix}\right\}$$

Eq. 11-63

The parameters that are needed to fully define this joint pdf are the two mean values $E\{\xi_1\}=\mu_1$ and $E\{\xi_2\}=\mu_2$, the two variances $E\{\xi_1^2\}-\mu_1^2=\sigma_1^2$ and $E\{\xi_2^2\}-\mu_2^2=\sigma_2^2$ and the covariances $E\{\xi_1\xi_2\}-\mu_1\mu_2=\sigma_{12}=\sigma_{21}$.

The matrix containing the variances and covariances is in fact called the *covariance matrix* $\mathbf{\Sigma}$.

The whole formula can be cast in a much more compact form as follows:

$$f_{\xi_1\xi_2}(x_1, x_2) = \frac{1}{2\pi\sqrt{\det\{\boldsymbol{\Sigma}\}}} \exp\left\{-\frac{1}{2}\left(\mathbf{x} - \boldsymbol{\mu}\right)^{\mathrm{T}} \cdot \boldsymbol{\Sigma}^{-1} \cdot \left(\mathbf{x} - \boldsymbol{\mu}\right)\right\}$$

Eq. 11-64

where:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

11.5.8.1 Important: correlation coefficient and statistical independence of jointly Gaussian random variables

We mentioned in Sect. 11.5.7 that the circumstance that the correlation coefficient of two RVs is zero is no guarantee that they are independent. An important special case in this respect is represented by *jointly Gaussian* RVs. For this kind of RVs, bi-implication occurs:

given two <u>jointly</u> <u>Gaussian</u> RVs ξ_1 *and* ξ_2*, then:*

$$\begin{Bmatrix} \text{statistical} \\ \text{independence} \end{Bmatrix} \Longleftrightarrow \left\{ \rho_{\xi_1 \xi_2} = 0 \right\}$$

Eq. 11-65

This can be easily seen as follows. We start from the definition of correlation coefficient:

$$\rho_{\xi_1 \xi_2} = \frac{\sigma_{\xi_1 \xi_2}}{\sigma_{\xi_1} \sigma_{\xi_2}} = \frac{\sigma_{12}}{\sigma_1 \sigma_2} = \frac{\sigma_{21}}{\sigma_1 \sigma_2}$$

Note that if the covariance $\sigma_{12} = \sigma_{21}$ is zero, then the correlation coefficient $\rho = \rho_{12} = \rho_{21}$ is zero too (see Eq. 11-62). Then the covariance matrix becomes diagonal and its inverse is just:

$$\mathbf{\Sigma}^{-1} = \begin{bmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{bmatrix}$$

As a result, with obvious calculations, the following result is found:

$$f_{\xi_1\xi_2}(x_1, x_2) =$$

$$= \frac{1}{2\pi \cdot \sqrt{\sigma_1^2 \sigma_2^2}} \exp\left\{ -\frac{1}{2} \left(\begin{bmatrix} (x_1-\mu_1) & (x_2-\mu_2) \end{bmatrix} \begin{bmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{bmatrix} \begin{bmatrix} (x_1-\mu_1) \\ (x_2-\mu_2) \end{bmatrix} \right) \right\}$$

$$= \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left\{ -\frac{1}{2\sigma_1^2} \left(x_1 - \mu_1\right)^2 \right\} \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left\{ -\frac{1}{2\sigma_2^2} \left(x_2 - \mu_2\right)^2 \right\} =$$

$$= f_{\xi_1}(x_1) \cdot f_{\xi_2}(x_2)$$

Clearly, the joint pdf factorizes into the two distinct marginals, which guarantees statistiscal independence of the two random variables ξ_1 and ξ_2 .

Therefore, it is proved that if $\sigma_{12} = \sigma_{21} = 0$ or, equivalently, $\rho_{12} = \rho_{21} = \rho = 0$, then this implies the statistical independence of the two jointly-Gaussian RVs ξ_1 and ξ_2.

Optional

It is possible to re-write the pdf of two jointly Gaussian random variables making the role of the correlation coefficient explicit.

We first remark that the invers of the covariance matrix can be written, in general, as:

$$
\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{21} & \sigma_1^2 \end{bmatrix} \frac{1}{\sigma_1^2\sigma_2^2 - \sigma_{12}\sigma_{21}} =
$$

$$
= \begin{bmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{21} & \sigma_1^2 \end{bmatrix} \frac{1}{\sigma_1^2\sigma_2^2} \frac{1}{1 - \frac{\sigma_{12}\sigma_{21}}{\sigma_1^2\sigma_2^2}} = \begin{bmatrix} \frac{1}{\sigma_1^2} & -\frac{\sigma_{12}}{\sigma_1^2\sigma_2^2} \\ -\frac{\sigma_{21}}{\sigma_1^2\sigma_2^2} & \frac{1}{\sigma_2^2} \end{bmatrix} \frac{1}{1-\rho^2}
$$

$$
= \begin{bmatrix} \frac{1}{\sigma_1} & 0 \\ 0 & \frac{1}{\sigma_2} \end{bmatrix} \frac{\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}}{1-\rho^2} \begin{bmatrix} \frac{1}{\sigma_1} & 0 \\ 0 & \frac{1}{\sigma_2} \end{bmatrix}
$$

Eq. 11-66

Substituting the above result into Eq. 11-64, the following result is found:

$$f_{\xi_1\xi_2}(x_1,x_2) =$$

$$= \frac{1}{2\pi\cdot\sqrt{\sigma_1^2\sigma_2^2-\sigma_{12}\sigma_{21}}}\exp\left\{-\frac{1}{2}\begin{bmatrix}(x_1-\mu_1) & (x_2-\mu_2)\end{bmatrix}\begin{bmatrix}\frac{1}{\sigma_1^2} & -\frac{\sigma_{12}}{\sigma_1^2\sigma_2^2}\\ -\frac{\sigma_{21}}{\sigma_1^2\sigma_2^2} & \frac{1}{\sigma_2^2}\end{bmatrix}\frac{1}{1-\rho^2}\begin{bmatrix}(x_1-\mu_1)\\(x_2-\mu_2)\end{bmatrix}\right\}$$

$$= \frac{1}{2\pi\cdot\sqrt{\sigma_1^2\sigma_2^2\left(1-\rho^2\right)}}\exp\left\{-\frac{1}{2}\begin{bmatrix}\frac{(x_1-\mu_1)}{\sigma_1} & \frac{(x_2-\mu_2)}{\sigma_2}\end{bmatrix}\begin{bmatrix}\frac{1}{\sigma_1} & 0\\ 0 & \frac{1}{\sigma_2}\end{bmatrix}\frac{\begin{bmatrix}1 & -\rho\\ -\rho & 1\end{bmatrix}}{1-\rho^2}\begin{bmatrix}\frac{1}{\sigma_1} & 0\\ 0 & \frac{1}{\sigma_2}\end{bmatrix}\begin{bmatrix}\frac{(x_1-\mu_1)}{\sigma_1}\\ \frac{(x_2-\mu_2)}{\sigma_2}\end{bmatrix}\right\}$$

$$= \frac{1}{2\pi\cdot\sqrt{\sigma_1^2\sigma_2^2\left(1-\rho^2\right)}}\exp\left\{-\frac{1}{2\left(1-\rho^2\right)}\left[\frac{\left(x_1-\mu_1\right)^2}{\sigma_1^2}+\frac{\left(x_2-\mu_2\right)^2}{\sigma_2^2}-2\rho\frac{\left(x_1-\mu_1\right)}{\sigma_1}\frac{\left(x_2-\mu_2\right)}{\sigma_2}\right]\right\}$$

Eq. 11-67

It is then easy to see that if $\rho = 0$ then immediately the joint pdf factorizes into the two marginals.

Finaly, even if $\rho \neq 0$ and the two RVs are statistically dependent, their marginals remain the same, for any value of ρ:

$$f_{\xi_1}(x_1) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left\{-\frac{1}{2\sigma_1^2}\left(x_1 - \mu_1\right)^2\right\}$$

$$f_{\xi_2}(x_2) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left\{-\frac{1}{2\sigma_2^2}\left(x_2 - \mu_2\right)^2\right\}$$

Simply, when $\rho \neq 0$ their joint pdf is not the product of the marginals.

End of optional material

11.5.8.2 Multiple jointly Gaussian RVs

Finally, note that the joint Gaussian pdf can be easily generalized from two RVs to an arbitrary number N of RVs. Its expression differs very little from Eq. 11-64:

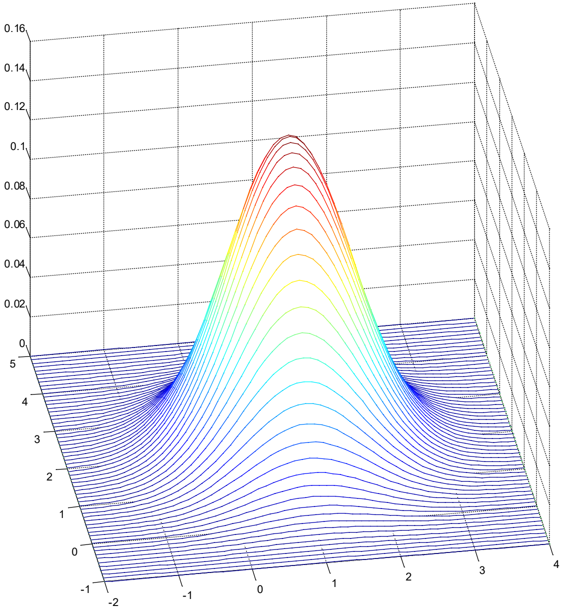
$$f_{\xi_1\xi_2\ldots\xi_N}(x_1, x_2, \ldots, x_N) = f_{\boldsymbol{\xi}}(\mathbf{x}) =$$

$$\frac{1}{(2\pi)^{N/2}\sqrt{\det\{\boldsymbol{\Sigma}\}}}\exp\left\{-\frac{1}{2}\left(\mathbf{x}-\boldsymbol{\mu}\right)^{\mathrm{T}}\cdot\boldsymbol{\Sigma}^{-1}\cdot\left(\mathbf{x}-\boldsymbol{\mu}\right)\right\}$$

Eq. 11-68

where:

$$\boldsymbol{\xi} = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_N \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_N \end{bmatrix} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \ldots & \sigma_{1N} \\ \sigma_{21} & \sigma_2^2 & & \\ \vdots & & \ddots & \\ \sigma_{N1} & & & \sigma_N^2 \end{bmatrix}$$


0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
5
4
3
2
1
0
-1
-2
-1
0
1
2
3
4

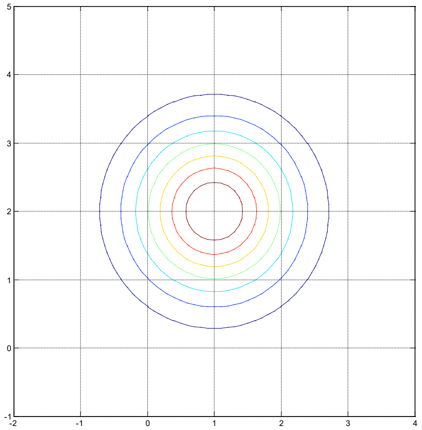
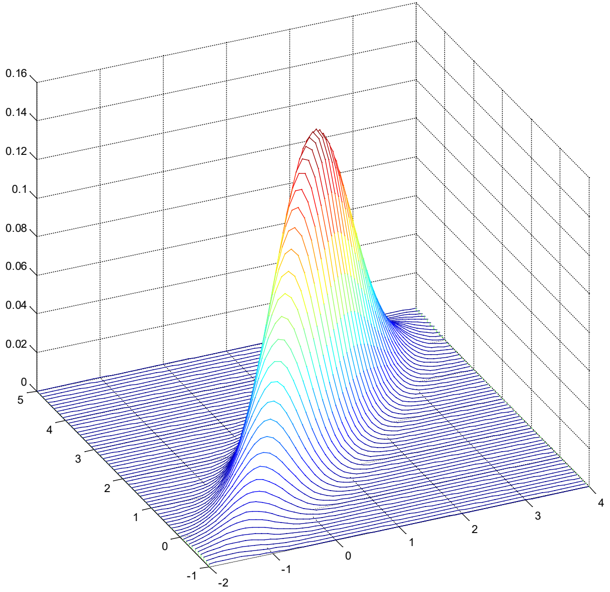


Figure 11-1: 3D and contour plot of the joint pdf of two <u>independent</u> Gaussian random variables, with means 1 and 2 and unit variance.



0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
5
4
3
2
1
0
-1
-2
-1
0
1
2
3
4

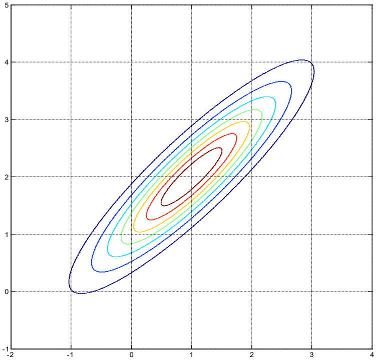


Figure 11-2: 3D and contour plot of the joint pdf of two dependent Gaussian random variables, with means 1 and 2, unit variance and correlation coefficient $\rho = 0.9$.

11.5.8.3 sums of RV's

A situation that arises quite frequently in practical situations is the linear combination of two or more RV's:

$$\rho = \sum_{n=1}^{N} a_n \xi_n$$

Eq. 11-69

where the ξ_n 's are RV's which can be, in general, dependent on one another, and the a_n 's are numbers. The result of such sum is a new RV ρ, for which the following general result holds.

Given a linear combination of generic RV's:

$$\rho = \sum_{n=1}^{N} a_n \xi_n$$

the mean value of ρ *is:*

$$\mu_\rho = \sum_{n=1}^{N} a_n \mu_{\xi_n}$$

Eq. 11-70

and its variance is:

$$\sigma_\rho^2 = \sum_{m=1}^{N} \sum_{n=1}^{N} a_m a_n \sigma_{\xi_m \xi_n}$$

Eq. 11-71

where:

$$\mu_{\xi_n} = E\left\{\xi_n\right\}$$

$$\sigma_{\xi_m\xi_n} = E\left\{\xi_m\xi_n\right\} - \mu_{\xi_m}\mu_{\xi_n}, \qquad m \neq n$$
$$\sigma_{\xi_n\xi_n} = E\left\{\xi_n^2\right\} - \mu_{\xi_n}^2 = \sigma_{\xi_n}^2, \qquad m = n$$

In the case when the RV's in the summation are statistically independent the variance becomes simply:

$$\sigma_\rho^2 = \sum_{n=1}^{N} a_n^2 \sigma_{\xi_n}^2$$

Eq. 11-72

These results are quite simple to prove. For the mean:

$$\mu_\rho = E_\rho\left\{\rho\right\} = E\left\{\sum_{n=1}^{N} a_n\xi_n\right\} = \sum_{n=1}^{N} a_n E_{\xi_n}\left\{\xi_n\right\} = \sum_{n=1}^{N} a_n\mu_{\xi_n}$$

For the variance:

$$\sigma_\rho^2 = E_\rho\left\{\left(\rho-\mu_\rho\right)^2\right\}$$
$$= E\left\{\left(\sum_{n=1}^{N} a_n\xi_n - \sum_{n=1}^{N} a_n\mu_{\xi_n}\right)^2\right\} = E\left\{\left(\sum_{n=1}^{N} a_n\left(\xi_n-\mu_{\xi_n}\right)\right)^2\right\}$$
$$= E\left\{\sum_{m=1}^{N} a_m\left(\xi_m-\mu_{\xi_m}\right)\sum_{n=1}^{N} a_n\left(\xi_n-\mu_{\xi_n}\right)\right\}$$
$$= \sum_{m=1}^{N} a_m\sum_{n=1}^{N} a_n E_{\xi_m,\xi_n}\left\{\left(\xi_m-\mu_{\xi_m}\right)\left(\xi_n-\mu_{\xi_n}\right)\right\} = \sum_{m=1}^{N}\ \sum_{n=1}^{N} a_m a_n \sigma_{\xi_m\xi_n}$$

The next question that one could ask is: what is the pdf of ρ? An important theorem provides the answer.

Given a linear combination of <u>independent</u> RV's:

$$\rho = \sum_{n=1}^{N} \xi_n$$

the pdf of their sum ρ *is given by:*

$$f_\rho(r) = f_{\xi_1}(r) * f_{\xi_2}(r) * f_{\xi_3}(r) * \ldots * f_{\xi_N}(r)$$

Eq. 11-73

In other words, the pdf of the sum of many independent RV's is the *convolution* of their pdf's.

Optional: The result Eq. 11-73 is important per se, as it provides the way for finding the resulting pdf. However, it acquires even more significance, since the convolution product has very special properties, which have a very substantial impact on RV theory.

One of these key properties is the following:

given two Gaussian functions, their convolution is still a Gaussian function.

The proof is immediate under Fourier transform. We cast it in the notation of RV theory and so we write the two Gaussian functions as $f_{\xi_1}(x)$ and $f_{\xi_2}(x)$, with respective mean and variance $\mu_{\xi_1}, \sigma_{\xi_1}$ and $\mu_{\xi_2}, \sigma_{\xi_2}$. However, for the purpose of this specific proof, these should be understood as simply generic Gaussian functions. We then have:

$$F\left\{f_{\xi_1}(x)\right\} = e^{2\pi^2 f^2 \sigma_{\xi_1}^2} e^{-j2\pi f \mu_{\xi_1}}$$

$$F\left\{f_{\xi_2}(x)\right\} = e^{2\pi^2 f^2 \sigma_{\xi_2}^2} e^{-j2\pi f \mu_{\xi_2}}$$

Remembering the standard properties of the Fourier transform:

$$F\left\{f_{\xi_1}(x) * f_{\xi_1}(x)\right\} = F\left\{f_{\xi_1}(x)\right\} \cdot F\left\{f_{\xi_1}(x)\right\}$$

$$= e^{2\pi^2 f^2 \sigma_{\xi_1}^2} e^{-j2\pi f \mu_{\xi_1}} \cdot e^{2\pi^2 f^2 \sigma_{\xi_2}^2} e^{-j2\pi f \mu_{\xi_2}}$$

$$= e^{2\pi^2 f^2 \left(\sigma_{\xi_1}^2 + \sigma_{\xi_2}^2\right)} e^{-j2\pi f \left(\mu_{\xi_1} + \mu_{\xi_2}\right)}$$

Finally, applying inverse Fourier transform:

$$f_{\xi_1}(x) * f_{\xi_1}(x) = F^{-1}\left\{e^{2\pi^2 f^2 \left(\sigma_{\xi_1}^2 + \sigma_{\xi_2}^2\right)} e^{-j2\pi f \left(\mu_{\xi_1} + \mu_{\xi_2}\right)}\right\}$$

$$= \frac{1}{\sqrt{2\pi\left(\sigma_{\xi_1}^2 + \sigma_{\xi_2}^2\right)}} e^{-\frac{\left(x - \left[\mu_{\xi_1} + \mu_{\xi_2}\right]\right)^2}{2\left[\sigma_{\xi_1}^2 + \sigma_{\xi_2}^2\right]}}$$

The result is clearly again a Gaussian function. In this particular notation, it is also clear that it is a Gaussian pdf. The new mean is the sum of the means and the new variance is the sum of the variances, in agreement with Eq. 11-70 and Eq. 11-72.

Once this property is proven for the convolution of two Gaussian pdf's, it is obviously proven for the cascaded convolution of any number of Gaussian pdf's. In, fact, we can generalize and find the property shown in the next section.

End of optional material.

11.5.8.4 sums of Gaussian random variables

A very important property of Gaussian random variables is the following:

given a linear combination of *<u>independent Gaussian</u>* *RV's:*

$$\rho = \sum_{n=1}^{N} a_n \xi_n$$

their sum ρ *is itself a Gaussian RV, with mean and variance:*

$$\mu_\rho = \sum_{n=1}^{N} a_n \mu_{\xi_n} , \qquad \sigma_\rho^2 = \sum_{n=1}^{N} a_n^2 \sigma_{\xi_n}^2$$

Eq. 11-74

Quite remarkably, it can be proved that if the Gaussian RV's that are summed together are *not independent*, then the resulting random variable ρ *is still Gaussian,* though the variance of the sum ρ is expressed by the more general formula Eq. 11-71. We can in fact state:

given a linear combination of <u>*dependent or independent*</u> <u>*Gaussian*</u> *RV's:*

$$\rho = \sum_{n=1}^{N} a_n \xi_n$$

their sum ρ is itself a Gaussian RV with mean and variance:

$$\mu_\rho = \sum_{n=1}^{N} a_n \mu_{\xi_n} , \qquad \sigma_\rho^2 = \sum_{n=1}^{N} \sum_{n=1}^{N} a_n a_m \sigma_{\xi_n \xi_m}$$

Eq. 11-75

Note that Eq. 11-74 for the variance is found from Eq. 11-75 by assuming statistical independence, which causes all $\sigma_{\xi_n \xi_m} = 0, \quad \text{for } n \neq m$.

Note also that in general, given a linear combination of RV's, whose pdf is *not* Gaussian, then the pdf of their sum has a *different* pdf than that of any of the summed RVs. So it is just for Gaussian random variables that their sum is still a Gaussian RV.

11.5.8.5 Optional: The Central Limit Theorem

An important and famous theorem, called the "Central Limit Theorem", shows that the Gaussian distribution spontaneously emerges whenever several independent random events are combined. This in turn justifies its ubiquity in physics, engineering and many other disciplines.

There are many forms of this theorem. The simplest one is:

Let $\left\{\xi_n\right\}_{n=1}^{N}$, *be a set of statistically independent RV's. Assume that they are zero-mean, that is* $\mu_{\xi_n} = 0$, *and that they are equally-distributed, with variance* $\sigma_\xi^2 < \infty$. *Assume that their distribution is any non-delta-like pdf.*

Then, the RV:

$$\rho = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \xi_n$$

tends to take on a Gaussian distribution, for $N \to \infty$, *with* $\mu_\rho = 0$ *and* $\sigma_\rho^2 = \sigma_\xi^2$.

There are even more powerful versions of the central limit theorem, in which the assumption that the RV's are equally distributed is relaxed. However, other conditions have to be met for convergence to Gaussian, so we omit the discussion.

The proof of the Central Limit Theorem actually hinges on two results. One was already given, as Eq. 11-73, which we re-write here:

$$f_\rho(r) = f_{\xi_1}(r) * f_{\xi_2}(r) * f_{\xi_3}(r) * \ldots * f_{\xi_N}(r)$$

The second result, which we do not prove, can be stated as:

given a function $f(r)$ *the result of the repeated convolution of such function with itself tends to become a Gaussian function, provided that the "variance" obtained by interpreting* $f(r)$ *as a pdf is finite.*

Therefore, when carrying out the repeated convolution Eq. 11-73, the result tends to become a Gaussian distribution. In fact, convergence to Gaussian is typically very fast for well-behaved distributions, such as uniformly-distributed RV's. "Good" Gaussian-distributed RV's can be generated by summing together between 10 and 20 uniformly-distributed RV's.

11.5.9 PDF's of RV transformations

As mentioned, in general, a new RV can be obtained as a function of another one, i.e.:

$$\eta = g(\xi)$$

We showed that the "transfer of expectation" property of the expectation operator makes it easy to calculate the moments or other averages of η based on the knowledge of the statistical properties of ξ.

A different problem is that of calculating the pdf of η. This is in general possible, with different degrees of difficulty depending on the features of the transformation g.

Here we provide the result for the simple case:

$$\eta = a\xi + b$$

with a and b constants. Then, given the pdf of ξ, $f_\xi(x)$, the pdf of η is:

$$f_\eta(y) = \frac{1}{|a|} f_\xi\left(\frac{y-b}{a}\right)$$

Start of optional material

In this simple case, the proof can be found as follows.

We first look at the case $a > 0$.

We univoquely relate an interval $[y_1, y_2]$ over the y axis with an interval over the x axis $[x_1, x_2]$ through the function $y = g(x) = ax + b$. We get:

$$\left[y_1, y_2\right] = \left[ax_1 + b, ax_2 + b\right]$$

where the corresponding values for x_1, x_2 must have values:

$$x_1 = \frac{y_1 - b}{a} \quad , x_2 = \frac{y_2 - b}{a}$$

This means that we have the correspondence through g of the intervals:

$$\left[x_1, x_2\right] \overset{g}{\leftrightarrow} \left[y_1, y_2\right]$$

Then, since the intervals correspond, it must be:

$$P\left(\left\{\eta \in [y_1, y_2]\right\}\right) = P\left(\left\{\xi \in [x_1, x_2]\right\}\right)$$

Eq. 11-76

The right-hand side of Eq. 11-76 is calculated as:

$$P\left(\left\{\xi \in [x_1, x_2]\right\}\right) = \int_{x_1}^{x_1} f_\xi(x)dx$$

$$= \int_{\frac{y_1-b}{a}}^{\frac{y_2-b}{a}} f_\xi(x)dx = \int_{y_1}^{y_2} \frac{1}{a} f_\xi\left(\frac{y-b}{a}\right)dx$$

Eq. 11-77

where in the last passage we used the change-of-variable $x = \dfrac{y-b}{a}$. The left-hand side of Eq. 11-76 is:

$$P\left(\left\{\eta \in \left[y_1, y_2\right]\right\}\right) = \int_{y_1}^{y_1} f_\eta\left(y\right) dy$$

Eq. 11-78

Comparing Eq. 11-77 with Eq. 11-78, it is evident that for their right-hand-sides to have the same value it must be:

$$f_\eta\left(y\right) = \frac{1}{a} f_\xi\left(\frac{y-b}{a}\right)$$

Redoing the same calculations with a negative value of a produces the similar result:

$$f_\eta\left(y\right) = \frac{1}{\left|a\right|} f_\xi\left(\frac{y-b}{a}\right)$$

So, the result for both positive and negative a can be unified as:

$$f_\eta(y) = \frac{1}{|a|} f_\xi\left(\frac{y-b}{a}\right)$$

Eq. 11-79

$\boxed{\text{On your own}}$: based on similar reasoning, prove that if $y = g(x)$ is generic but has either everywhere-positive or everywhere-negative derivative (in other words, it is a strictly monotone function, and therefore also invertible everywhere), then the transformation:

$$\eta = g(\xi)$$

leads to:

$$f_{\eta}(y) = \frac{1}{\left|g'\left(g^{-1}(y)\right)\right|} f_{\xi}\left(g^{-1}(y)\right)$$

Eq. 11-80

where $g'(x) = \dfrac{d\,g(x)}{dx}$. Show that Eq. 11-80 produces exactly Eq. 11-79 when assuming $g(x) = ay + b$.

End of optional material.

11.6 Problems

11.6.1 problem (optional)

Consider a deck of 52 cards. As customary, there are 13 cards of hearts, diamonds, clubs and spades, respectively. We assign value 1 to aces, value 11 to jacks, 12 to queens and 13 to kings. All other cards have value identical to their numbers (2 to 10).

1. Define the space of the events of the random experiment consisting of extracting one card at random from the deck.
2. Characterize it probabilistically.
3. Defining: $A = \{\text{the card seed is diamonds}\}$, $B = \{\text{the card value is less than } 3\}$, find $P(A)$, $P(B)$, $P(A \cap B)$
4. Are A and B mutually exclusive? Are A and B statistically independent?

Answers

$P(A) = 1/4$, $P(B) = 2/13$

$P(A \cap B) = 1/26$

A and B are not mutually exclusive.

A and B are statistically independent.

11.6.2 problem (optional)

Consider the random experiment consisting of throwing two dice and multiplying their results.

5. Find the space of events and characterize the probability of each event.
6. Defining: $A =$ {result being odd}, $B =$ {the result being less than or at most equal to 9}, find $P(A)$, $P(B)$, $P(A \cap B)$
7. Are A and B statistically independent? Are A and B mutually exclusive?

Answers

The space of event consists of:

$$\{1,2,3,4,5,6,8,9,10,12,14,15,16,18,20,24,25,30,36\}$$

where the labels are the results of the product of the two dice.

The respective probabilities are:

$$\left\{\frac{1}{36},\frac{1}{18},\frac{1}{18},\frac{1}{12},\frac{1}{18},\frac{1}{9},\frac{1}{18},\frac{1}{36},\frac{1}{18},\right.$$

$$\left.\frac{1}{9},\frac{1}{18},\frac{1}{36},\frac{1}{18},\frac{1}{18},\frac{1}{18},\frac{1}{36},\frac{1}{18},\frac{1}{36}\right\}$$

Also:

$$P(A)=\frac{1}{4},\ P(B)=\frac{17}{36}\ ,\ P(A\cap B)=\frac{1}{6}$$

The events A and B are NOT mutually exclusive, since $A\cap B=\{1,3,5,9\}\neq\varnothing$

They are NOT statistically independent, since:

$$P(A\,|\,B)=\frac{6}{17}\neq P(A)=\frac{1}{4}$$

11.6.3 problem

Describe how you would approximate the probability density function of a random variable, based on the results of a finite number N of random experiments.

What formula would you use?

Show that the integral of the approximation is always 1 over $\Omega = \mathbb{R}$, independently of N

11.6.4 problem

We assume that ξ is defined as:

$$\xi = \sin(\varphi)$$

We then assume that φ is a uniformly-distributed random variable in the range $[-\pi/2, \pi/2]$. Find the mean value, mean square value, variance, root mean square value and standard deviation of ξ.

Answers

$$\mu_\xi = E_\xi\{\xi\} = 0$$

$$E\{\xi^2\} = \frac{1}{2}$$

$$\sigma_\xi^2 = \frac{1}{2}$$

$$\mathrm{RMS}_\xi = \sqrt{E\{\xi^2\}} = \frac{1}{\sqrt{2}}$$

$$\sigma_\xi = \sqrt{\sigma_\xi^2} = \frac{1}{\sqrt{2}}$$

11.6.5 problem

We assume that ξ is defined as:

$$\xi = \exp\left(-a \cdot \eta\right)$$

We then assume that η is a uniformly-distributed random variable in the range $[0,1]$. Find the mean value, mean square value, variance, root mean square value and standard deviation of ξ.

Answers

$$\mu_\xi = E_\xi\{\xi\} = \frac{1-e^{-a}}{a}$$

$$E\{\xi^2\} = \frac{1-e^{-2a}}{2a}$$

$$\sigma_{\xi}^{2}=\frac{1-e^{-2a}}{2a}-\left(\frac{1-e^{-a}}{a}\right)^{2}$$

$$\mathrm{RMS}_{\xi}=\sqrt{E\left\{\xi^{2}\right\}}=\sqrt{\frac{1-e^{-2a}}{2a}}$$

$$\sigma_{\xi}=\sqrt{\sigma_{\xi}^{2}}=\sqrt{\frac{1-e^{-2a}}{2a}-\left(\frac{1-e^{-a}}{a}\right)^{2}}$$

11.6.6 problem

We assume that ξ is defined as:

$$\xi=\eta^{2}$$

We then assume that η is a Gaussian-distributed RV with $\mu_{\eta}=0$ and $\sigma_{\eta}^{2}=1/2$

Find the mean value of ξ.

Answer

$$\mu_\xi = \frac{1}{2}$$

Hint: use the transfer of expectation property and then look at the integral that comes out and you could interpret it.

11.6.7 problem

Consider the following RV's:

- ξ is zero-mean and Gaussian, with variance σ_ξ^2
- ρ is zero-mean and Gaussian, with variance σ_ρ^2, and is independent of ξ
- η is defined as: $\eta = a\xi + b\rho$
- γ is defined as: $\gamma = a\xi$

Find the correlation coefficient $\rho_{\eta\gamma}$.

Check that indeed the general property $\left|\rho_{\eta\gamma}\right| \le 1$ is verified.

Discuss the result.

Answer

$$\rho_{\eta\gamma} = \sqrt{\frac{a^2\sigma_\xi^2}{a^2\sigma_\xi^2 + b^2\sigma_\rho^2}}$$

11.6.8 problem

With reference to the example 11.5.4.5:

- calculate $\rho_{\xi\eta}$ and discuss the result
- calculate:

$$P\left(\left\{\xi \in \left[0, \tfrac{1}{2}\right]\right\} \cap \left\{\eta \in \left[0, \tfrac{1}{2}\right]\right\}\right)$$

Answers

- $\rho_{\xi\eta} = \dfrac{1}{\sqrt{2}}$

  hint: substitute $\eta = \xi + \rho$ and use the transfer of expectation property; then optionally you can also check that the same result is obtained by carrying out a direct calculation using the joint pdf:

  $f_{\xi\eta}(x, y) = \Pi_1(x)\Pi_1(y - x)$

- $P\left(\left\{\xi \in \left[0, \tfrac{1}{2}\right]\right\} \cap \left\{\eta \in \left[0, \tfrac{1}{2}\right]\right\}\right) = \dfrac{1}{4}$

Solution

Theory says that given any region of the plane $\mathcal{D} \subset \mathbb{R}^2$, then the probability of $P\left(\left\{(\xi, \eta) \in \mathcal{D}\right\}\right) = \iint_{\mathcal{D}} f_{\xi\eta}(x, y)\,dxdy$.

According to the above formula:

$$P\left(\left\{\xi\in\left[0,\tfrac{1}{2}\right]\right\}\cap\left\{\eta\in\left[0,\tfrac{1}{2}\right]\right\}\right)=P\left(\left\{(\xi,\eta)\in\mathcal{D}=\left[0,\tfrac{1}{2}\right]\times\left[0,\tfrac{1}{2}\right]\right\}\right)$$

$$=\int_0^{1/2}\int_0^{1/2} f_{\xi\eta}(x,y)\,dxdy=\int_0^{1/2}\int_0^{1/2}\Pi_1(x)\Pi_1(y-x)\,dxdy$$

The order of integration is completely arbitrary. If we integrate first in y, which seems the easier option, we get for the inner integral:

$$q(x)=\int_0^{1/2}\Pi_1(y-x)\,dy$$

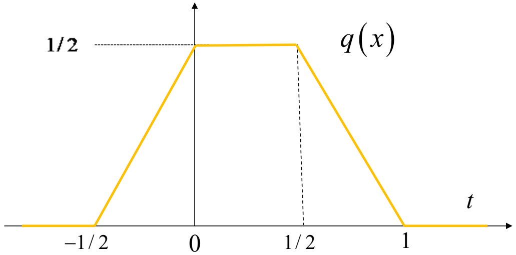
The support in y of the integrand function is:

$$\underset{y}{\operatorname{supp}}\left\{\Pi_1(y-x)\right\}=\left[-\frac{1}{2}+x,\frac{1}{2}+x\right]$$

Looking at the intersection of this support with the integration interval, which is $\left[0,\frac{1}{2}\right]$, we can calculate the integral easily:

$$
\begin{cases}
x < -1/2 \quad , \quad q(x) = 0(x) \\
-1/2 < x < 0 \quad , \quad q(x) = \int\limits_{0}^{x+1/2} 1(y)\,dy = x + 1/2 \\
0 < x < 1/2 \quad , \quad q(x) = \int\limits_{0}^{1/2} 1(y)\,dy = 1/2 \cdot 1(x) \\
1/2 < x < 1 \quad , \quad q(x) = \int\limits_{x-1/2}^{1/2} 1(y)\,dy = 1 - x \\
x > 1 \quad , \quad q(x) = 0(x)
\end{cases}
$$

The plot of the resulting function is:



We then have to calculate the remaining outer integral:

$$\int_0^{1/2} \Pi_1(x) q(x)\, dx = \int_0^{1/2} q(x)\, dx = \frac{1}{2} \int_0^{1/2} 1(x)\, dx = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

Note that calculating the inner integral $q(x)$ for all values of x was not strictly necessary. One could have noticed that the outer integral limited the relevant values of x to just $\left[0, \frac{1}{2}\right]$. So $q(x)$ could have been calculated only for those values of

x, resulting in a much shorter solution.