Rules of the game

- this file is a draft
- maybe more definitions at the start e.g. a random variable, a pdf. Or start with couple words about preliminaries[1]
- probably all the "we" should be changed to "I" (there is only one narrator).
- the video should be from 15 to 20 minutes long. Given this, some parts may be removed (like subsection 2.5)
- maybe make a list of formulas somewhere on the screen in order for it to not get lost
- does really "функция постоянной тени касания" translates to "the function of constant tangency shadow"?

Содержание

[1] the latter is arguably better because explaining expectation and variance isn't fun and otherwise inevitable

1 motivation

Hi![2]. This video is about the intuition behind the density function of a normal distribution.

Normal distribution lies in the center of all statistics. It is recognized in various real world occasions, which allows it to be widely used in assumptions about data. Moreover, it has many fantastic properties that allow deep mathematical insights.

Normal distribution is the leading character in mathematical statistics, so that virtually anyone who took a course in probability theory recalls its name, as well as the formula that characterizes it — the proaBbility density function of a normal distribution[3]. Unfortunately, many fellow students think of it as a given and doesn't get to know why does it have this particular form. I mean, why on earth is there an exponential and not some other function? And what does it have to do with π? [4]. In reality, these questions have a reasonable answers that require barely any knowledge of probability theory.

Today me and a particle named Bb will walk you through the derivation of this formula, so by the end of the video you will know the origins of such fundamental object[5]. And also some fun facts from visual calculus that will help us along the way.

2 assumptions

This story is about the motion of a particle[6], and not just one particle, but a whole swarm of particles![7].

Meet Bb[8] — a lonely oxygen particle mindlessly floating in space at different speeds. The speeds are dictated by random fluctuations and collisions of Bb and other particles in a swarm. Let's measure her speed in all three directions[9].

These vectors represent the speed of a particle, and they are going to be our random variables V_1, V_2, V_3. Together they form a random vector V[10]. You may be tempted to say that these distributions are equal, but this is not what we're going to do now. At this stage we try to be as general as possible and try to avoid excessive assumptions. So no, these velocities does not have the same distribution, at least not yet. We've only assumed that each of these random variables has a distribution.

[2] rotationg 2-dim normal pdf plot

[3] big wide expression for a pdf with μ and σ^2

[4] these questions among other (make sure they are concisely written) show up

[5] some kind of animations not to be static

[6] linear motion of 1 particle in a box with it bouncing like a dvd logo on a grid (!)

[7] other particles appear and it becomes the 3-dim brownian motion

[8] Bb is colored pink

[9] motion of Bb slows down (?), arrows start to appear

[10] $V = \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix}$ attached to Bb

2.1 [Rot]

What natural properties should these distributions have? I mean, like the property that V_1 should not always equal to some number v_0. [11] In a real world all the particles are colliding with each

x	v_0
$\mathbb{P}\{V_1 = x\}$	1

other, thus changing their speeds very fast, so no one can say for sure what velocity does Bb have at a given time moment.

Also, observe that the velocity can be any real number. You may ask, what units of measure do we use, and the answer may surprise you! The units can be anything you like! In fact, we will later fix a useful unit of measure in order for a certain assumption to hold. But at this moment, it feels safe to say that irrespective of units, Bb can achieve any given speed, positive being forward and vice versa.

Continuing the description of particles' world perception, Bb is kind of indifferent to orientation. She doesn't care if she moves up or down, to the west or to the east. Her small world doesn't have the notion of direction. Hence, if we rotate our axes by, for example, 90 degrees counterclockwise[12], changing east to south, north to east; the distribution of each random variable will not change.

90 degrees are not special — any angle rotation will not change the distribution.

This assumption is called "rotational assumption"[13]. It can be formulated as follows: "rotating the distribution by any angle about the origin leaves the distribution unchanged".

Great, the first assumption is present. Let's move to the next one.

2.2 [Proj]

Now we're going to study the projections of Bb's trajectory onto the orthogonal spaces[14].

These arrows are the three vectors that were pointing from Bb all this time. Together they form the random vector V.

If an observer knows the realization of V_1[15], what information about other projections should it give to him? It seems natural to assume that it must give no additional information whatsoever. In statistical terms, this means that V_1, V_2 and V_3 are pairwise independent. Traditionally, independence of two random variables is defined as follows: "the probability of the variables taking values simultaneously is equal to the product of the probabilities"[16]. But there is an equivalent formulation that involves conditional probabilities, in other words expressions of a type "probability of X being small

[11] a table with probability mass appears

[12] particles disappear leaving only the axis. Directions: positive x-axis becomes east, positive y-axis is north, positive z-axis is "up", negative -//-. The copy of the exis rotates around z-axis. They rotate around z-axis (original axis stay in place).

[13] [Rot] appears

[14] the borders of spaces appear. They are squares but the perspective makes them parallelograms. Different colors. Projections appear (vectors are still moving slowly alongside Bb?)

[15] projection of V_1 is highlighted

[16] alongside speaking these words, the formula $\mathbb{P}\{X = x,\ Y = y\} = \mathbb{P}\{X = x\} \cdot \mathbb{P}\{Y = y\}$ is highlighted from left to right

x if Y is equal to y"[17]. If we apply Bayes' formula, this formulation sounds like "the conditional probabilities is equal to the probability as if there is no information about Y"[18].

The name of this assumption is called "projection assumption"[19].

2.3 Unit variance

Now, let's revisit the discussion about the units of measure. As we said, velocity can be any real number. For example, if the particles moves fast in the forward direction, the value of V_1 is going to be positive and large and vice versa.

But what is "large" when we talk about a velocity of a particle? Is the value 1000 large? Perhaps from a human's point of view, 1000 meters in second is large, and 10 centimetres in a minute is small[20].

If you think about it, the world of air particles is indifferent to the units of measure used by humans, so Bb cannot care less about what metric is used to track her velocity. At the same time, it is necessary for us to choose some unit. Otherwise we would have no way of formalizing the problem in probabilistic terms. Then why not choose the units of measure that are most useful for us?

But what is useful? In our case turns out we want the variance of each random variable to be equal to 1.[21] To 1 unit of whatever you name it, squared (because remember that variance is measured in the square units).

This way we've found a way to fix the units and introduced a useful assumption at the same time!

2.4 Existence of a differentiable pdf

Okay. Now the only thing left to assume is to say that each random variable has a differentiable probability density function. [22]

The following are the graphs of the joint pdf[23] that takes two values and outputs a positive number that we mark at the vertical axis[24].

In fact, if we stick to our initial problem of figuring out the distribution of the speed in a 3-d space, we would need to come up with a function of three variables, v_1, v_2 and v_3 that also outputs one positive number. This function cannot even be visualized intuitively: for that we would have needed a 4-d space, which is more then can be asked from life. As we'll see later, we're going to break this multidimensional case into the smallest parts, that is, into 1-d densities. They not only can be easily visualized, but are also a subject to basic school calculus analysis.

Actually, the fact we're trying to prove can still be proven without mentioning differentiable densities, but this way it becomes rather technical, lacking clarity and elegance.

[17] $\mathbb{P}\{X = x \mid Y = y\}$ is also highlighted from left to right

[18] $\mathbb{P}\{X = x \mid Y = y\} = \mathbb{P}\{X = x\}$ is highlighted

[19] [Proj] appears

[20] animation for 1000 meters/sec a particle flies over Manhattan (included graphics of a map with the right perspective?), for 10 cm/min a particle alongside a snail

[21] [Var]: $\operatorname{Var} V_1 = \operatorname{Var} V_2 = \operatorname{Var} V_3 = 1$ appears

[22] [pdf] appears, graphs of various pdfs with a support of $\mathbb{R}^2$ appear

[23] $f_V(v_1, v_2)$ appears

[24] oZ is highlighted

2.5 Why assumptions?

Before moving on, let's talk a bit about one fundamental question: "why do people introduce assumptions in the first place?"

This question may seem strange, but try to imagine yourself playing a role of a scientist who picks assumptions for a certain model. The choices you make are going to affect how flexible versus constrained you're going to be in the process of proving facts about this model. In the process of studying some branch of mathematics, you start to understand what facts gives you the highest flexibility. For example, continuity in calculus, full rank in linear algebra, perfect competitiveness in economics, and...[25] independence in probability theory! In fact, this assumption is going to be so useful that our problem will be transformed from a 4-dimensional one to a 2-dimensional. This is a great motivation to include an assumption into your model.

Also, independence is a special property in a sense that there are countless ways for random variables to be dependent, each of them requiring special consideration, and there is only one way the variables can be independent. A quote by famous russian novelist Leo Tolstoy comes to mind[26]: "Happy families are all alike; every unhappy family is unhappy in its own way"[27].

To finish up the discussion, we're going to point out that physical models like ours usually benefit well from assumptions that are present in the world around us. Take the [Rot] assumption: the fact that air particles do not concentrate in one part of a room is a direct corollary of it. And sure enough, you'll see how useful this assumption is in just a minute!

But for the sake of good visualizations and intuitive connections, let's go down by one dimension[28]

3 Finding the form of a pdf

Let's denote the pdf of the vector V by a bold symbol $\boldsymbol{f}(v_1, v_2)$.[29]. Recall that we've assumed it exists[30]

Let's begin utilizing the assumptions we've made.[31] Since we can rotate our coordinate axis as desired, let's make a 90-degree rotation[32]

This way, by the [rot] assumption $\boldsymbol{f}(v_1, v_2) = \boldsymbol{f}(v_2, v_2')$, but $v_2' = -v_1$, so $\boldsymbol{f}(v_1, v_2) = \boldsymbol{f}(v_2, -v_1)$. The same logic implies that we can add minus signs everywhere, so let's remember this fact.

Now, notice that by [Proj] assumption we have $\boldsymbol{f}(v_1, v_2) = \gamma(v_1)\gamma(v_2)$, where gammas are the marginal densities of V_1 and V_2. This can be intuitively thought of as the continuous analog of the

[25] pause

[26] the portrait of Tolstoy appear in the bottom

[27] the quote types itself from the left and the year 1877 appears

[28] 3-d brownian motion is starting to be looked from above, thus turning into 2-d brownian motion. The third vector disappears, only two axis left

[29] $pdf_V(v_1, v_2) := \boldsymbol{f}(v_1, v_2)$

[30] the = sign between the LHS and the RHS is highlighted and [pdf] appears in overset (indication that we've used an assumption)

[31] brownian motion disappears, two orthogonal (classic) coordinate axis appear named v_1 and v_2

[32] axis are duplicated in manim and the copy rotates 90-degrees counterclockwise. Now $v_2' = (-1, 0)$ and $v_1' = (1, 0) = v_2$

independence definition[33].

Now we can manipulate this[34] result and get

$$\gamma(v_1)\gamma(v_2) = \gamma(v_2)\gamma(-v_1),$$

or just $\gamma(v_1) = \gamma(-v_1)$.

This is great news, because now we are ready to make a really helpful conceptual observation. Since $\boldsymbol{f}(v_1, v_2)$ does not depend on direction and treats all angles equally, it must depend on the length! What we mean by that is the following: if you pick any point inside a petri dish that is a given amount of distance away from the origin, all those point must have equal probability density, and the only thing that affects the value of the density is its distance from the origin.

In other words, $\boldsymbol{f}(v_1, v_2) = h(v_1^2 + v_2^2)$. This is because the length of a vector in two dimensions is the square root of the sum of the squared of its coordinates[35] and we can abandon the square root due to the fact that it is an increasing function.

This suits well with the result we've just established:[36] since the function γ is even, it must have the following property: $\gamma(v_1) = g(v_1^2)$. Recall that Taylor series expansions of even functions include only terms raised to an even power.[37]

How we arrive at the first significant result, which is the combination of the last two:

$$\boldsymbol{f}(v_1, v_2) = g(v_1^2) \cdot g(v_2^2) = h(v_1^2 + v_2^2).$$

Wait, this is very similar to the property of an exponential! Recall that e^{a+b} reduces down to $e^a \cdot e^b$! If only the functions h and g could be the same...

In fact, one is just another times a constant, and now you'll see why! Notice that since this equality holds for all numbers v_1, v_2, we can set v_2 to be zero and get

$$h(v_1^2 + 0) = h(v_1^2) = g(0) \cdot g(v_1^2).$$

The function g at zero is just some number, let's call it "g underscript zero"[38]. Similarly[39],

$$h(0 + v_2^2) = h(v_2^2) = g(v_2^2) \cdot g_0$$

But notice that we can multiply these two equations and get

$$g_0^2 \cdot g(v_1^2) \cdot g(v_2^2) = h(v_1^2) \cdot h(v_2^2) = g_0^2 \cdot h(v_1^2 + v_2^2)$$

This is when h (which is just an another name for $\boldsymbol{f}$) really starts to look like an exponential function. In fact, this property is sometimes used as a definition, and the fact that it is equivalent

[33] $\mathbb{P}\{A, B\} = \mathbb{P}\{A\}\mathbb{P}\{B\}$ appears

[34] highlighted

[35] the formula $||x|| = \sqrt{x_1^2 + x_2^2}$

[36] $\gamma(v_1) = \gamma(-v_1)$ highlighted

[37] Taylor expansions appear and the terms are highlighted

[38] g_0 appears

[39] the original formula $h(v_1^2 + v_2^2)$ is copied and transformed two times

with other definitions generally involves transforming this into a differential equation. That being said, let's do it the other way around. We're going to take the logarithm

$$\ln h(v_1^2 + v_2^2) = \ln h(v_1^2) + \ln h(v_2^2) + \mathrm{const}(v_1, v_2)$$

Let's rename things a bit

$$l(a + b) = l(a) + l(b) + \mathrm{const}(a, b)$$

and take the derivative with respect to a

$$\frac{d}{da} l(a + b) = \frac{d}{da} l(a) + 0 + 0$$

which needs to be true for all a, b.

Wait a minute, this means that l is just a linear function!

$$l(a) = k_1 a + k_2 = \ln h(v_1^2, v_2^2)$$

this tells us that the function h itself of an exponential family!

$$h(v_1^2) = e^{k_1 v_1^2 + k_2} =: h_0 \cdot e^{k_1 v_1^2}$$

All of this brings us to the conclusion:

$$\boldsymbol{f}(v_1, v_2) = h(v_1^2 + v_2^2) = h_0 \exp\{k_1(v_1^2 + v_2^2)\}$$

The only thing left is to find the values of h_0 and k_1.

4 Finding the constants

4.1 Finding h_0

4.1.1 Strategy

Let's start with h_0. It's going to be very helpful to draw a picture. Luckily, the plot turns out to be very pretty: it's going to have many symmetries because v_1 and v_2 can be interchanged and negated without affecting the value of $\boldsymbol{f}$.

Also, we can confidently say that the coefficient in front of x^2 is negative. Let's just put a minus sign in front of it and remember that k is positive. If it wasn't, then in the limit our function $\boldsymbol{f}$ would be positive infinity, which violates the fact that a volume under the density function must integrate to one.

So, our plot looks something like this bell-shaped surface[40]. We see that $\boldsymbol{f}(0, 0) = h_0$, which is the highest point of a bell, and it's going to determine how stretched is our surface. To get the value of h_0, we're going to integrate this function, so that the volume is always equal to one.[41].

[40] 3d bell curve is drawn
[41] the mass is highlighted

You may be familiar with integration by slicing the plane with vertical lines[42]. But we're going to do it the other way: our slices are going to be horizontal, in other words we're slicing the height and adding up the corresponding circular areas[43].

To calculate the total volume we consider small changes in height[44], multiply them by the area of a slice[45], and then sum up. The limit of this procedure is precisely what it means to take a three dimensional integral.

We notice that the distance from the origin is the radius of some circle[46], which makes the radius squared equal to the sum of the squares of coordinates[47].

This allows us to write[48]

$$h(R^2) = h_0 \cdot e^{-kR^2},$$

which makes

$$R^2(h) = \frac{\ln(h) - \ln(h_0)}{-k}.$$

Remember that we wanted to integrate over the vetrical "height" axis, so the left hand side[49] of the formula is precisely our variable of integration.

Finally, let's put everything together. We want to have an integral over the areas to be equal to one

$$1 = \int_0^{h_0} S(h)dh,$$

where the area is

$$S(h) = \pi R^2(h) = \pi \cdot \frac{\ln(h) - \ln(h_0)}{-k},$$

which makes

$$1 = \int_0^{h_0} \pi \cdot \frac{\ln(h) - \ln(h_0)}{-k} dh.$$

Now let's rearrange the terms

$$\frac{-k}{\pi} = \int_0^{h_0} \ln\left(\frac{u}{h_0}\right) du$$

Finally, remember that the differential dh[50] acts like a derivative, so we can write it as follows[51]

$$\frac{-kh_0}{\pi} = -\int_{u=0}^{u=h_0} \ln\left(\frac{u}{h_0}\right) d\left(\frac{u}{h_0}\right).$$

[42] a grid and vertical slices are drawn
[43] horizontal slices are drawn
[44] z-axis is highlighted with many tiny ticks
[45] a slice is highlighted
[46] the formula with square roots appears
[47] square roots get removed
[48] the corresponding variables are transformed with the whole formula inplace
[49] highlighted
[50] highlighted
[51] du is now $d(u)$ and then $h_0 \cdot d(u/h_0)$

For the sake of elegance, let's introduce a new variable $t = u/h_0$, so that when $u = 0$ we have $t = 0/h_0 = 0$ and when $u = h_0$ our new variable t must be $h_0/h_0 = 1$[52]. In the end, we get

$$\frac{-k}{\pi h_0} = \int_0^1 \ln t dt.$$

This looks like a trivial integral that we know how to evaluate. But let's have some fun and find it using no calculus at all!

4.1.2 Constant shadow

To do so, we are going to encounter three visually appealing facts. First, symmetry: let's graph the function we're integrating over.[53] Now shade the region we want to know the area of[54]. The fact that $\ln t$ and e^t are inverses of each other is fairly well-known, but I'm going to remind you of a useful visual representation. Two functions are inverses if they are symmetric around this bisector line[55]. Notice that this splits the 90 degree angle between axis in two equal parts.[56]

The mirrored graph looks familiar! It's always positive, it intersects the y-axis at $y = 1$, it grows fact. It must be the exponential! And it surely it, which you can prove as an exercise. Now, since the functions are equal, we must have some other way to formulate the integral. If we look closely enough, we notice that

$$\int_0^1 \ln t dt = \int_{-\infty}^0 e^t dt.$$

This feels even more easy, but we're just getting started.

So, everybody knows an exponential function. But what if I told you that it didn't have this name back in the end of the 18'th century, when such figures as Euler and Lagrange were laying the ground for analysis. These days the exponential function was called "a function of constant tangency shadow"[57].

What a strange name! Surely, it must be motivated by some property, and it indeed does.

Let's draw a tangent line at $t = 0$[58], or better the left half of it[59], and remove everything below the vertical x-axis[60]. We got a segment, which forms a right triangle ABC[61]. Let's gather the facts we know about it.

Firstly, it is indeed a right triangle since the angle ACB[62] is formed by two axis, which are perpendicular by default. Secondly, since we've decided to draw the tangent segment at $t = 0$, the value of our function $\exp\{t\}$ at zero is the length of a segment BC[63].

[52] animation: table ?
[53] the graph of $\ln t$ appears
[54] the region from 0 to 1 is shown
[55] a line $y = x$ is shown
[56] the graph is mirrored along a bicestor line
[57] this naming is shown with big bold letters
[58] drawn
[59] the right part is moving up with lowering the opacity
[60] the lower part is moving down with lowering the opacity
[61] the letters appear clockwise
[62] indicated somehow
[63] $BC = e^0 = 1$ is shown

Okay, BC is of length 1, what do we also know? Since our hypotenuse is a segment of a tangent line, we must be able to figure out its slope, which is a characteristic of this angle[64]. Recall that the derivative at some point is the slope of a tangent line at this point, and the slope is the tangent of an angle. If you don't have a solid ground recalling this, remember that a tangent is the quotient of a vertical change with the horizontal change, given you rotate the triangle appropriately. And the slope is precisely the same thing: the vertical change divided by the horizontal one.

Having said that, we know that the derivative must equal to BC divided by AC[65], and this is easy because the derivative of an exponential is the same exact exponential![66]. This all brings us to the conclusion that AC must also be equal to one!

Indeed, if we no not specify t explicitly and give it a name t_0, which can be any number on a x-axis[67], then the length of BC is going to be e^{t_0}, the tangent of alpha is also e^{t_0}, which makes AC of unit length once again.

Convict yourself that this is true by inspecting the length of BC segment for different values of t[68]. Surprisingly, while the lengths of other two sides of the triangle change, we see that AC stays constant.

That's why many great mathematicians of the 18'th century referred to this fact as a foundational property of an exponential function.

Okay, cool. But how on earth can we utilize this fact?

4.1.3 Mamikon

This is the place when we move to the third, and final trick, named after an Armenian mathematician Mamikon Mnatsakanian. This fact is actually a theorem from a little-known branch of calculus called visual calculus.

Visual calculus is described by Wikipedia as follows: "*Many problems that would otherwise seem quite difficult yield to the method with hardly a line of calculation, often reminiscent of a proof without words*".[69]

[64]angle BAC is highlighted

[65]in the center, the formula

$$\frac{BC}{AC}$$

appears, then gets changed to

$$\frac{BC}{AC} = \frac{1}{AC}$$

[66]the previous formula

$$\frac{BC}{AC} = \frac{1}{AC}$$

is now

$$\frac{BC}{AC} = \frac{1}{AC} = \frac{d}{dx}e^t\bigg|_{t=0} = e^t\Big|_{t=0} = 1$$

[67]while this paragraph is read, the whole construction is sliding horizontally

[68]while this paragraph is read, the whole construction is sliding horizontally

[69]a Wikipedia page quote is shown

Truly a magical description, isn't it? All of this beauty lies on a reasonably intuitive visual fact: suppose you have a smooth path, alongside which you slide a tangent segment of fixed length[70].

Now consider the area that is traced by the endpoint of a tangent segment[71], and suppose you want to find the area is this segment.

It turns out that if we consider all the infinitesimal segments for each time moment[72] and move them to start from one single point, then their total area is going to be equal to the one we've traced.

In fact, we can allow the segment to change its length during the procedure.

To highlight the practical utility of this fact, we're going to prove an interesting property of the exponential function, and then apply Mamikon's trick to it.[73]

To illustrate the application of Mamikon's method, we're gonna make it shine in all of its glory! Consider the setup we've just build: let this tangent segment[74] be one of Mamikon's segments, and then study the area that it makes while cnanging the coordinate of intersection from zero to minus infinity[75].

Cool! We see that it is a portion of the area we want to find[76], and what's really nice is that the other portion is just an area of a right triangle: the same exact triangle formed by a tangent line at t and two axis[77]. It's area is just two perpendicular sides multiplied, then divided by 2[78].

Now back to the Mamikon area. We want to somehow utilize the "constant tangency shadow" property by dragging all infinitesimal segment to a certain position where they all start from a single point. Now is a good place to pause the video and try to see it yourself[79]. We won't need any new constructions, so it's not going to require much creativity, just some observation skill[80].

It may come as a surprise, but the following configuration turns out to be especially convenient[81]. It may seem a little slippery, but everything is valid[82]: the top segment, which is our good old AB, stays inplace[83]; the lowest one, which is infinitely far to the left, can be assumed having almost zero slope, is going to become AC in the limit, and all the segments in the middle are going to take a place somewhere between the two.

[70] a simple yet not trivial 2d object (not a circle) is shown (on a grid ?), then a half length tangent line is sliding over it

[71] the area is highlighted

[72] one by one (with ascending pace), single segments are rapidly increasing in opacity (highlights). The segments are chosen randomly

[73] maybe the frame is squeezing by a factor of 8 approx and stays in the LR corner while the part with constant shadow is on the screen

[74] highlighted

[75] the animation is done

[76] the integral pops up for a second and hides one again

[77] the triangle is somehow highlighted (maybe a rectangle is outlined)

[78] the animation of getting $AB \cdot BC/2 = 1/2$ is shown

[79] while the sentence is read, all segments are shifted to various locations, maybe even including the right one

[80] pause 4 sec, maybe everything turns black & white for this period of time

[81] the segments are shifted to fill the triangle

[82] reversed

[83] AB is highlighted

You may ask, what about these overlaps and extra parts outside of triangle ABC? Well, our illustrations are inevitably going to include those, but in the limit these are going to become infinitesimal, thus we can neglect these.

I hope you're convinced in the validity of this procedure, and at the moment you're sitting amazed by how some arbitrary strange looking shape suddenly became another, trivial to analyze shape with the same area.

With the help of Mamikon, we conclude that the total value of the desired integral is twice the area of ABC, namely one half times two equals one![84]

Of course, this integral does not really require any trick in solving it. And it's always good to check the creative solution with a more standard one. In our case, we could have remembered that we need to find an antiderivative of a function inside the integral and then substitute in the values from the limits.

Actually, let's do so: an antiderivative is a function, which derivative is the initial function. In our case the initial one is e^t, which points us towards the same exponential, since the derivative doesn't change it. Indeed, the derivative of e^t is e^t, which means our final answer would be e to the power of the upper limit–zero–minus e to the power negative infinity. The first part is easy: e to the power of zero is one, and the second one is not as hard as it may seem. Understand it as follows: what happens if we raise a number to the negative power? School lessons tell us that it is the same as taking the reciprocal[85], so when we're approaching an exponent of minus infinity, we're just considering the value of a fraction one divided by e, which is raised to bigger and bigger values. Your intuition should indicate towards it approaching zero in the limit, and as a matter of fact, it is indeed zero, which makes the answer one minus zero equals to one – the same thing as we got earlier with Mamikon's method!

Great, so what's next? Weren't we talking about the density function of the normal distribution? Yes, now we need to recall the equation which led us to discussing integrals [86].

$$\frac{-k}{\pi h_0} = \int_0^1 \ln t dt.$$

Now we know that the right hand side of this is minus one, which gives us $h_0 = k/\pi$. Remember what we had earlier:

$$\boldsymbol{f}(v_1, v_2) = h_0 \exp\{-k(v_1^2 + v_2^2)\}$$

so now we have k divided by π instead of h_0, and let's split the function into the product of exponentials:

[84]formula $\int_0^1 \ln t dt = \int_{-\infty}^0 e^t dt = 1$ appears

[85]the formula

$$a^{-b} = \frac{1}{a^b}$$

is shown

[86]in the first place

$$\boldsymbol{f}(v_1, v_2) = \sqrt{\frac{k}{\pi}} \exp\{-kv_1^2\} \cdot \sqrt{\frac{k}{\pi}} \exp\{-kv_2^2\}$$

these two are probability density functions for one dimensional distributions: this is ensured by our [Proj] assumption in a way that the speed in some dimension does not give any information about the speed in other ones.

$$\boldsymbol{f}(v_1, v_2) = f(v_1)f(v_2)$$

Well, since f, which is not in bold font, is a density, we may go on exploring its properties, which hopefully will lead to finding the value of k – the only parameter left.

5 Finding k

This is a good place to stop and ponder because there are assumptions we haven't yet used. We see that the unit variance assumption[87] is still available to us.

Let's utilize it for a final sprint towards the closed form of f! We need to start from the defining formula for variance:

$$\operatorname{Var} X = \mathbb{E}\left[(X - \mathbb{E}X)^2\right],$$

which can be though of intuitively as the expected squared difference between a random variable and its mean. If a random variable is not too random, which allows it to take values in some small region of space with high probability, for example this normal distribution[88], then we can expect it to deviate from the mean–which is zero in this case–very rarely. This way the variance is going to be small.

On the other hand, consider a random variable with a very wide probability density: there are no specific range of values which is distribution takes with considerably higher probability then any other. In this case[89], we see that the outcomes are likely to deviate from the mean. In this case the distribution has large variance.

With this intuition in mind, remember why we've decided to introduce this variance assumption in the first place. No matter whether the distribution of interest has high variance or not, we are the ones who choose the units of measure. Bb is in some sense senseless of the units: they are just a human abstraction to make the language of physics formal. This is why we're doing this act of "normalization": we fix our units of measure by saying that the variance of all V's is one, when expressed in these units.

Mathematically, this allows us to manipulate the equality about variances and get the value of k from it. But first, we need to figure out the expectation of V_i, since it is present in the definition of variance.

At this point, there are many ways to do so. Perhaps, the most standard would be to integrate the density. But we're going to use a physics-based argument about the meaning of the expectation.

[87] $\operatorname{Var} V_i = 1 \; \forall i \in \{1, 2, 3\}$

[88] a normally distributed r.v. with zero mean and small variance is shown. The samples are being drawn

[89] the pdf disappears, only the mean and continuously drawing samples remain

Remember at the beginning we were defining these random variables V_1 up to V_3 as velocities of some particle. Then the expectation of these r.v.'s is the expected velocity at some point is space. What will happen if these expected values are non-zero? Intuitively, then all the air particles in the room would eventually fly to one side of the room. Mathematically, we would have a bias towards some directions, and this violates the [Rot] assumption: the distribution must be rationally symmetric.

We've just proven that the expectation of all V's is zero! Let's substitute it into the definition of variance and see what simplifies:

$$1 = \operatorname{Var} X = \mathbb{E}\left[(X-0)^2\right] = \mathbb{E}\left[X^2\right],$$

And this this we know how to get! It is just the following integral:

$$1 = \int_{-\infty}^{+\infty} x^2 \sqrt{\frac{k}{\pi}} \exp\{-kx^2\} dx$$

This is just a matter of being able to utilize integration by parts from this point. We're going to stop here, leaving the link to an amazing technique of integration by parts from a YouTuber BlackPenRedPen[90] and a written derivation without the voiceover[91]. In the end, you would get that k is just one divided by two, which leads to the final result being

$$f(v_1) = \sqrt{\frac{1}{2\pi}} \exp\left\{\frac{v_1^2}{2}\right\},$$

which is precisely the pdf of a standard normal distribution

[90] the link and the video in a mini box appears

[91] by a change of variables $t = \sqrt{k}x$, which means $dx = dt/\sqrt{k}$ and $-t^2 = -kx^2$. Also, the bounds of integration don't change, so we rewrite the integral

$$1 = \int_{-\infty}^{+\infty} \frac{t^2}{-k} \sqrt{\frac{k}{\pi}} e^{-t^2} \frac{dt}{\sqrt{k}} = \frac{1}{-k\sqrt{\pi}} \int_{-\infty}^{+\infty} t^2 e^{-t^2} dt = \frac{1}{k\sqrt{\pi}} \int_{-\infty}^{+\infty} \frac{t^2}{2t} d(e^{-t^2}).$$

Using integration by parts,

$$k\sqrt{\pi} = \frac{t}{2} e^{-t^2}\Big|_{t=0}^{t=+\infty} - \int_{-\infty}^{+\infty} e^{-t^2} d(t/2) = 0 - 0 - \frac{1}{2} \int_{-\infty}^{+\infty} e^{-t^2} d(t).$$

Since $f(t) = \sqrt{k/\pi} \exp\{-kt^2\}$ is the pdf, it must integrate to 1 for all values of k. In other words,

$$\int_{-\infty}^{+\infty} \sqrt{\frac{1}{\pi}} \cdot e^{-t^2} = 1,$$

or equivalently,

$$\frac{1}{2} \int_{-\infty}^{+\infty} e^{-t^2} = \frac{\sqrt{\pi}}{2}.$$

This gives the final result,

$$f(t) = \sqrt{\frac{1}{2\pi}} \exp\left\{\frac{t^2}{2}\right\},$$

6 exercises

If you was inspired by this derivation, I suggest you to try changing it a bit to get a more general normal distribution density with variance σ^2 and mean μ. Also, here are the references that you may like, which can extend your knowledge of this method.

7 conclusion

If fact, I haven't yet revealed the people after which this derivation is named: it is called Herschel Maxwell axioms, and if you want to learn more about it, we highly suggest watching a similar video on the 3blue1brown channel. In fact, Grant has recently uploaded several episodes dedicated to probability theory. They are truly top quality, so anyone can enjoy this tale of math.

Bye! Bb was very happy to get to meet you.[92]

[92]Hedgehog with red heart emoji