Chapter 1

Continuous Random Variable

> It all comes down to the density of the wood that makes every guitar different
>
> Robin Trower

What is continuum?

So far we have been working with discrete random variables. This chapter is devoted to continuous random variable and its properties. In particular, we will introduce the Normal random variable, the most widely used distribution in Statistics.

Keep in mind, a random variable is a *numerical value of something*. So far we analyzed discrete random variable, and now we will switch to continuous random variable, which in contrast to discrete, has infinitely many possible values even on a finite interval. In other respects the logic of analysis is similar.

For instance, if we randomly choose a person from population, his *height* would be a continuous random variable. It is random because we are talking about the height before we randomly select a person so his or her height is determined by chance. And it is continuous because height can be *any* number within some reasonable bounds. It is said that a person is 175 cm tall though in reality one is 175.3478 cm, another is 175.3476 cm and the number of possible outcomes within an interval 175,176 cm is thus infinite. We round off height for simplicity because such a precision does not matter, but in fact person's height is a continuum of numbers.

Other examples of continuous variables, which fit naturally into this category, are variables measuring time, distance, or temperature like:

- Tomorrow's temperature at 17 o'clock

- The speed of a car on road at a random moment

- Time to wait the bus

In such a case the probability that a random variable takes on a particular single value, like for instance, the probability that a car will go precisely 125.236 km/h is zero since it is trivially small. Also we regard as continuous some countable variables that are measured on such a scale, that probability of occurring some highly precise specific value is too small and meaningless, and thus assumed as 0. Example is monthly income. It is of no use to count probability income will be \$1502.05 or \$1502.06 and thus random variable is treated as continuous.

In such cases instead of occurrence specific values of a variable, we are interested in small intervals – given ranges. For example, we are interested in probability that monthly income is not less than \$1500, or the probability that the speed of car at a moment will be from 120km/h to 130km/h.

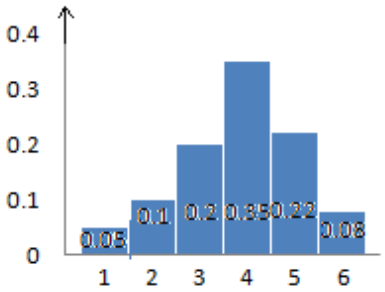
Probability distribution of continuous variable

To work with any random variable we need to know how it is distributed, formally its probability distribution. Probability distribution is the set of values a variable can take with the associated probabilities. How to find such probabilities of a continuous random variable?

To answer this question let's first go back to the discrete case and think how can probability distribution be shown on a graph. This will help you to apply the same logic for continuous case: to understand how probabilities look like graphically and then how to calculate them mathematically.

"ROOMS & HOUSES" If we randomly select a house the number of rooms X in this house would be a discrete random variable, and the height of rooms H – a continuous random variable. Consider probability distribution of X which is provided in table below. What is meant by, for example, probability $P(X = 3) = 0.20$? It tells us that 20% of all houses in that area have 3 rooms. As we mentioned, the probability can also be shown on a graph. It will look like an *area* of a column above value 3. This column contributes 20% of the total area of all columns. Every column (bar) is associated with a specific value and represents the corresponding area. All columns together form a total area of 1.

X	1	2	3	4	5	6
$P(X = x_i)$	0.05	0.10	0.20	0.35	0.22	0.08



In our example the width of a bar is equal to 1, therefore the area of a bar is equal to its height. It is not necessarily like that. Further, as you know the probability of an interval of values equals the sum of probabilities of each value within this interval. Graphically it looks like summation of column areas between the required values inclusively. For example, $P(2 \leqslant X \leqslant 4) = P(X = 2) + P(X = 3) + P(X = 4) = 0.1 + 0.2 + 0.35 = 0.65$. This interval contributes 65% of total area. So this graph shows probability as part of total area.

Probability = AREA

Now, think how can probability be shown graphically for continuous case? Discrete variable has finite number of possible values, on an interval, whereas continuous has infinite. Making equivalent graph would require drawing infinitely many bars. Moreover, the width of each bar would approach zero, since probabilities of specific values are tiny small. The bar will be just a line. Imagine this picture. If you connect the peaks of these imaginary zero-width bars you'll get a curve. This curve is what is known as $f(x)$ **probability density function**. But be careful, the pdf does *not* show probability. Remember, probability is not only the height, but the area, while pdf reflects only the height! So pdf values are not full probability. To get probability you have to multiply height by width.

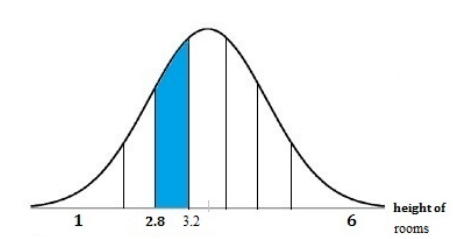
What is meant by width here? The width between values is actually zero. The probability is continuously accumulated. Instead of bars you will have lines of zero-width which you need to sum up. This will result in ... an integral! It does not contradict geometry though: the area of a single line is zero, while the area of an interval is positive.

Thus, we arrive to the formula of probability of a continuous random variable:

$$P(a \leqslant X \leqslant b) = \int_a^b f(x)dx$$

where $f(x)$ is the probability density function.

Now, consider the probability distribution of H. For example, the probability that the height of room in a randomly chosen house is approximately 3 m, or specifically is between 2.8 and 3.2 m is equal to $P(2.8 < H < 3.2) = 0.15$. That means 15% of all houses in that area have such a height of rooms. On pdf graph this will look like that. The probability 0.15 is equal to shaded area. All together it will look like that.

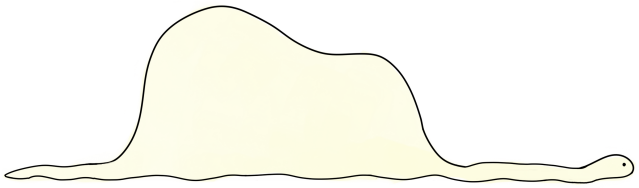


At the same time $P(H = 2.8) = 0$ or $P(H = 3) = 0$, etc, since such a high precision of height is negligible. The total probability distribution of H is represented by the total area under the pdf curve. Each fragment (interval) of interest corresponds to particular values along the horizontal axis.

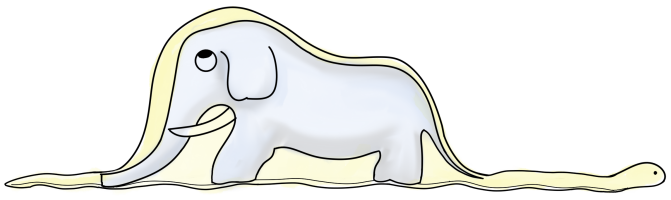
Note again, probabilities are "hidden" in the areas *under* the pdf and not in values of pdf itself. The values of pdf function do *not* have a probabilistic interpretation. Height of bar was coincided with probability in discrete case only because the width of bars was 1. So, pdf is a function that is required for further calculation of probability as an area under it. It is like only one dimension while probability is a two dimensions thing on this graph. To find probability we have to multiply height reflected by $f(x)$ by the width.

$f(x) \neq$ probability

Pdf may be of any form. It may look like that for instance.



It is only a contour, probability is hidden inside. Just like The Little Prince said, elephant is inside.



Grown-ups answered, "Why be scared of a hat?" My drawing was not a picture of a hat. It was a picture of boa constrictor digesting an elephant. Then I drew the inside of the boa constrictor, so the grown-ups could understand. They always need explanations.

The brief summary of comparison between discrete and continuous probability distributions is provided in the table below.

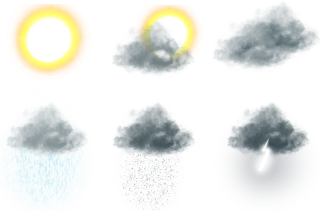
Property	DRV	CRV
Number of possible values	Finite	∞
Width of bar	$x_{n+1} - x_n$	tiny small
Probability of specific value	$P(X = x_i) = , 0 \leqslant c$	$P(X = x_i) = 0$
Probability of interval of values	$P(a \leqslant X \leqslant b) = \sum_i^n P(X = x_i)$, sum of bars	$P(a \leqslant X \leqslant b) = \int_a^b f(x)\,\mathrm{dx}$, integral

Notion of Probability Density

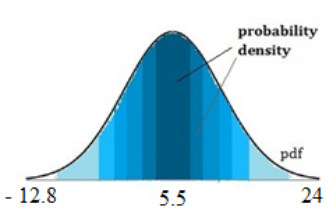
Let us get closer to the concept of density. Usually probabilities associated with values of CRV are called probability densities. The higher is the probability of a variable to be between some values, the higher is the density of observations on this interval. This is reflected by the bigger area (higher line of the graph as well).

The notion of probability density is exactly the same as population density. The higher density of some geographic areas is often denoted as darker inks on the maps. Another intuition could be made with the butter on bread. If hill is big that means there is a lot of butter (observations) in this place, so it has higher density.

"TOMORROW'S WEATHER" October in Moscow... ICEF student Masha is preparing going out tomorrow and planning what she could wear. She wonders what could be the daily temperature tomorrow?!



Based on previous observations temperature in October could vary from -12.8°to 24°*(Remark: Maximum and minimum observed during a century).* However, she feels that some *values* are *more likely* to occur. Based on past experience she knows that the random variable "daily temperature" *most probably* will be somewhere around 5.5°. This interval is denoted below with dark blue ink.



You see that the area around 5.5°is the biggest. That means the biggest portion of possible temperature lies between 4 and 5 degrees. Here line goes higher than in

other places. This means that density on this interval is higher making those numbers more probable to occur.

Properties of pdf:

- It is a function of x (depends only on one variable)

$$\text{pdf} = f(x)$$

- It is always non-negative for all values of x

$$f(x) \geqslant 0$$

- Total area under the pdf curve is always equal to 1

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

Cumulative Distribution Function

Cumulative probability is the *accumulated* probability up to some point. It is the probability that a random variable X *does not exceed* some value X. The notion of cumulative probability was also already introduced in Chapter 2. Now we will see it on continuous case, and here it is the function since probability is accumulating continuously. It is represented by **cumulative distribution function** cdf denoted as $F(x)$:

$$F(x) = P(X \leqslant x)$$

cdf is a non-decreasing function, taking values from 0 (for all values below minimum) up to 1 (for all values above or equal to maximum). Visual illustration is provided in the example below.

How are pdf and cdf connected?

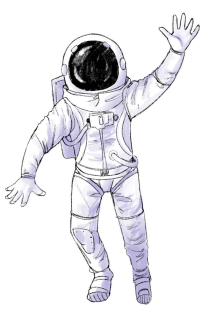
However, based on what we know about pdf, $P(X \leqslant x_a) = \int_{-\infty}^{x_a} f(x)dx$
So, $F(x_a) = \int_{-\infty}^{x_a} f(x)dx$
And we realize that the cumulative distribution function (cdf) and the probability density function (pdf) are connected in a very straightforward way:

$$f(x) = F'(x)$$

Cdf is a crocodile who ate pdf.

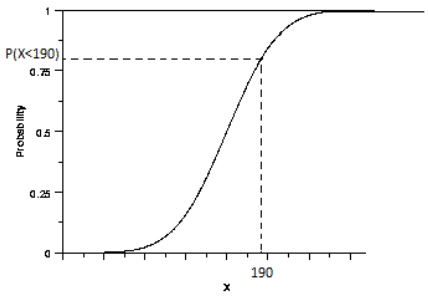
Recall from calculus that $\int_a^b f(x)dx = F(b) - F(a)$ for calculating interval probabilities

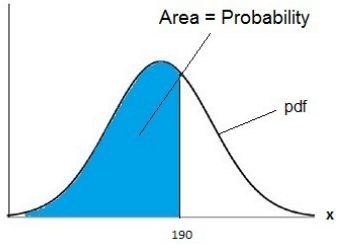
"ASTRONAUT HEIGHT" If we are randomly choosing a candidate to be astronaut, the height of this person would be a continuous random variable, denote it as X. The maximum acceptable



height for an astronaut is 190 cm. We do not care if a height of a person is exactly 175.87 or 175.85 cm, we only need it to be less than 190. So, chosing a candiadate we are interested in the probability of a random variable X not to exceed a specific value 190. Namely the probability that a person will suit, is the cumulative probability $P(X \leqslant 190)$. Assume it equals $P(X \leqslant 190) \approx 0.79$. How to denote this probability on a graph?

On the pdf it is the area under the pdf curve to the left of $X = 190$, and on cdf it is the value of cdf corresponding to $X = 190$:





Expectation and Variance of Continuous Random Variable

In Chapter 2 we've introduced notions of expected value and variance of a random variable. Those are two main *characteristics* of a distribution. Here these notions are as well applied to the continuous case.

Just as in discrete case the *expectation* and *variance* are the two main numerical characteristics of a distribution. The mean provides a *measure of the center of the probability distribution* while variance – a *measure of the spread* of the probability distribution around its center. The definitions, notations and meanings of these characteristics are also the same.

The *expected value* or the *mean* of a continuous random variable X is denoted as:

$$E(X) = \mu$$

The *variance* of a random variable X is denoted as:

$$Var(X) = \sigma_x^2$$

and is defined as the expectation of the squared discrepancy of the random variable from its mean:

$$\sigma_x^2 = E(X - \mu)^2 = E(X^2) - \mu^2$$

In the "Top secret information" below you can find the derivation of these formulas for calculating expectation and variance of a continuous random variable.

The mean and variance are two pieces of *numerical summary* about a probability distribution. Just like the passport details of a person: Name and Surname. Note, if two persons have the same name they are still two different persons. The same is applied to distributions. If means or variances of two variables are equal, it does not mean the two distributions are the same.

Now let us investigate one type of continuous random variable - normal random variable. It exist other types of continuous variables as well, like for example student distribution or chi-squared distribution studied later. However, normal distribution plays a particular role in Statistics.

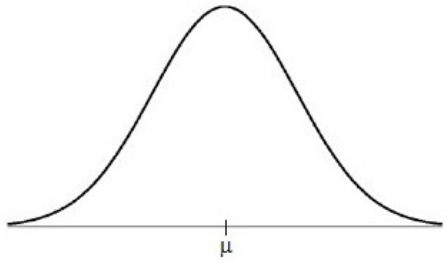
Normal Distribution

It happens that in life many variables are normally distributed, in other words follow normal probability distribution, this is why this distribution we are most often faced with. For instance, normal distribution is used to analyze deviation in sizes of manufactured items, or random errors of shooting. For other examples, it is also used in control of technological processes, to measure noises of radio-engineering devices, and many other things. In biology many variables tend to have approximately normal distribution including measure of size, length, height and weight of individuals. Many phenomena follow normal law. This is why Normal distribution is so popular and commonly used.

Probability Distribution of Normal random variable

As we've learned, the probability distribution for continuous variables (which is the collection of values with associated probabilities) is presented by the probability density function, pdf. It can also be presented by cdf, which together are the two ways of specifying the same thing, however the pdf is graphically richer so used more.

Normal pdf looks like this:

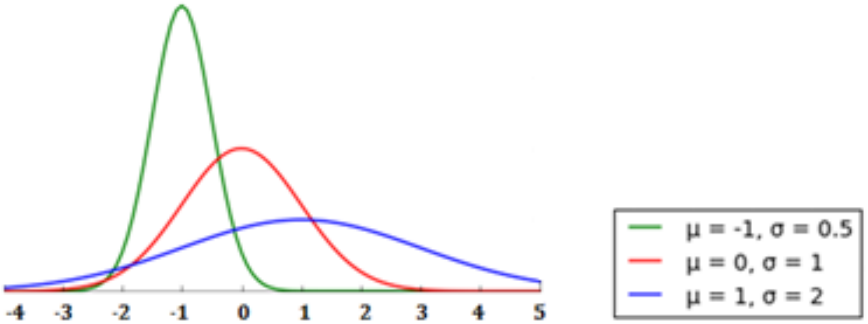


It is characterized by two parameters, mean and standard deviation. This is denoted $X \sim N(\mu, \sigma^2)$ which is translated as random variable X is distributed normally with mean μ and standard deviation σ. For instance, let X be a height of an adult man. Then $X \sim N(176, 6.5^2)$.

Properties of normal curve:

- It is bell shaped and symmetric
- It is centered around its mean
- It has high concentration of values in the center and low concentration in the "tails"
- It has infinite base namely it can take any value between $-\infty$ and $+\infty$

The distribution is solely defined by two characteristics – mean and standard deviation. Changing these characteristics will result in different distributions. You can check on a graph below how the form of normal distribution will change with different mean and standard deviation. As it is set by two parameters, changing them will change the form of curve.



The particular case is when $\mu = 0$ and $\sigma = 1$. This distribution is called **standard normal**.

How to find Probabilities for Normal Variable?

As it was explained, for a continuous random variable probability that it will be between certain values is found as an *area* under the pdf curve. This area can be calculated as integral of the probability density function:

$$P(a \leqslant X \leqslant b) = \int_a^b f(x)dx = F(b) - F(a)$$

Probability density function for a random variable $X \sim N(\mu, \sigma^2)$ is set by the following formula:

$$f(x) = \frac{1}{\sqrt{2}} e^{-\frac{(x-\sigma)^2}{2^2}}$$

At first glance it looks scaring. But, first check that its shape is completely determined by only 2 parameters – mean μ and standard deviation σ. Plugging particular values of μ and σ you will get a function depending only on x. Second you do not need to use this formula. It will be explained below how to find areas under normal curve using special tables or calculator.

Using Tables of Normal Distribution Check how to find normal probabilities. For example, let the variable be normal with 0 mean and standard deviation 1. As it was mentioned such a variable is accepted to be called standard normal and usually denoted by Z. Values of variable can be called scores.

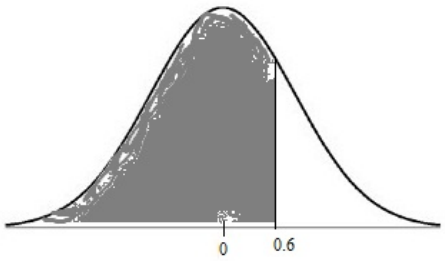
$$Z \sim N(0, 1^2)$$

Such probabilities are already calculated and are provided in special table. You will have this table on your exam too. The table provides probabilities that variable Z is *less* than some value z, namely $P(Z < z)$. It is the area under the normal curve to the *left* of this point. Probabilities are given in the bulk (body) of the table while corresponding values z are provided along the edges.

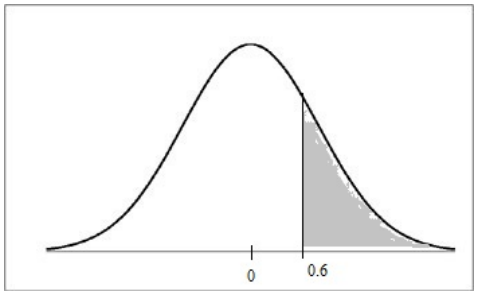
How the probabilities were calculated? Plug $\mu = 0$ and $\sigma = 1$ into general formula. You will get pdf of standard normal variable, namely $f(z) = \frac{1}{\sqrt{2}} e^{-\frac{z^2}{2}}$. Then take integral of this function on the required intervals, from $-\infty$ to z, and you will get the required probabilities.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	

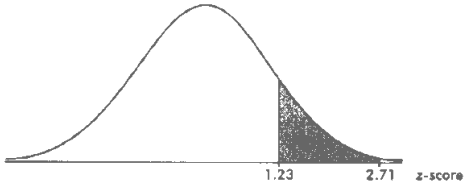
For example it shows, that to the left of a value $z = 0.6$ the area is equal to 0.7257. That means probability of Z to be less than 0.6 is 0.7257, $P(Z < 0.6) = 0.7257$. This can be also shown on a graph as:



Since the total area under the normal curve is, as for any pdf, 1, then the area to the right of $z = 0.6$ is equal to 1-0.7257 or 0.2743. The probability that Z is greater than 0.6 is 0.2743, $P(Z > 0.6) = 1 - 0.7257 = 0.2743$.



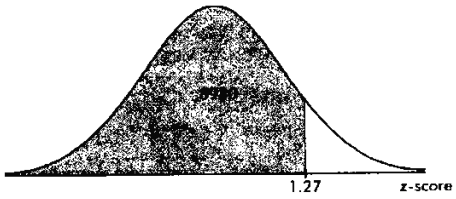
The area *between* two points can be found as subtraction of one area from another. For example, the area between the z-scores 1.23 and 2.71 is equal to the area to the left of *bigger* z value 2.71 (equals to 0.9966) minus the area to the left of smaller value 1.23 (equals to 0.8907) and is 0.1059.



$$P(1.23 < Z < 2.71) = P(Z < 2.71) - P(Z < 1.23) = 0.9966\text{-}0.8907 = 0.1059$$

Inverse problem

Given the value of probability we can use the same table to find the associated value of z. For example, to find the z that has an area of 0.8982 to the left of it, we search in table for the probability closest to 0.8982 and check to which z it corresponds. We found 0.8980 with the corresponding $z = 1.27$.



$$0.8982 = P(Z < z) \text{ thus } z = 1.27$$

If we are given probability to the right of some z value, we first need to subtract it from 1 in order to search in the table. For example, to find the z with an area of 0.72 to the right of it, we look for the probability 1-0.72 = 0.28 in the table. The probability 0.2810 corresponds to the z value -0.58.

$$0.72 = P(Z > z)$$

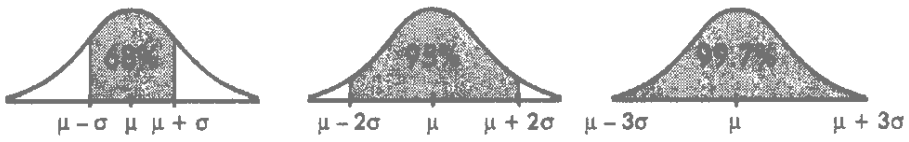
$$0.28 = 1 - P(Z > z) = P(Z < z) \text{ thus } z = -0.58$$

Always draw a graph of normal curve when solving problems on normal distribution! Sometimes it is the only way to avoid mistakes!

As you might notice probabilities given in the table $P(Z < z)$ are cumulative probabilities. So, the table provides you with values of cdf of standard normal random variable.

Another property of normal distribution is the **1, 2 and 3 sigmas rule**:

- approximately 68% of the observations falls within one standard deviation of the mean,
- approximately 95% of the observations falls within two standard deviations of the mean,
- approximately 99.7% of the observations falls within three standard deviations of the mean.



Transformation to standard normal variable

Standartization. We talked how to find probabilities for normal variable with 0 mean and 1 standard deviation. How to do it for all other distributions with other values of mu and sigma? Now it will come a little bit of magic. Fact is that any normal variable can be transformed into a standard normal variable. Let's check how it works.

To do that, mean should be moved into 0 and sigma to 1. Initially, X is centered at mu. How to move it to zero? Quite intuitively, by subtracting μ from each value we move the center into 0. The new normal variable is $X-\mu$. Its standard deviation is still sigma. So on average $X-\mu$ deviates from 0 by sigma. That's why division by sigma moves the standard deviation into 1.

Let's check that $\frac{X-\mu}{\sigma}$ is indeed $Z \sim N(0, 1^2)$.

First, it is a linear transformation of normal variable X, thus, is itself normal. Now let's check the parameters:

$$E(Z) = E(\frac{X-\mu}{\sigma}) = E(X\frac{1}{\sigma} - \frac{\mu}{\sigma}) = \frac{1}{\sigma}E(X) - \frac{\mu}{\sigma} = \frac{1}{\sigma}\mu - \frac{\mu}{\sigma} = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = \mathbf{0}$$

$$Var(Z) = Var(\frac{X-\mu}{\sigma}) = \frac{1}{\sigma^2}Var(X) - 0 = \frac{1}{\sigma^2}\sigma^2 = \mathbf{1}$$

So, it is true that $Z \sim N(0, 1^2)$.

Thus, it is true that any normal variable can be transformed into a standard normal variable by subtracting its mean μ and dividing by its standard deviation σ.

any normal random variable can be transformed to standard normal

$$Z = \frac{X-\mu}{\sigma}$$

So, knowing probabilities for standard normal variable we can apply them to any other normal variable. We can calculate integrals every time for every particular distribution (for particular μs and σs). But it is more convenient to calculate probabilities only for one distribution to which all others can be transformed.

We don't need to calculate integrals! Someone already did it for us!

Let us check how it works. For instance, let X be a height of a student. Assume $X \sim N(182, 4^2)$. Then, the probability to meet a student who is even taller than

190 cm is $P(X > 190)$. We proved that subtracting its mean μ and dividing by its standard deviation σ we will transform X to Z. Let us do that from both sides of inequality. From left side we will get Z. And from right side the particular value of z.

$$\frac{X - \mu}{\sigma} = Z$$

$$\frac{190 - 182}{4} = 2, \text{ where } \mu = 182 \text{ and } \sigma = 4$$

That means that the value 190 in the scale of X corresponds to the value 2 in the scale of Z. Like we transformed meters into cm. Since both variables are normal, they have the same normal distribution except for parameters. So we can switch from one distribution to another with linear transformation.

Namely the probability that X is greater than 190 is the same as the probability that Z is greater than 2.

$$P(X > 190) = P(\frac{X - \mu}{\sigma} > \frac{190 - 182}{4}) = P(Z > 2) = 1 - 0.9772 = 0.0228$$

How did we get 0.0228? From the table of normal probabilities.

Note that values of Z can be also thought of by how many standard deviations σ does particular value X exceed its population mean μ. For example let X be a height of a randomly taken student. It known that $X \sim N(182, 4^2)$. A boy from this population has height $X_1 = 190$ cm, which can be transformed into $Z_1 = \frac{190-182}{4} = 2$. This means that he is by 2 standard deviations taller than an average student in his population ($190 - 182 = 8$cm is equal to $2\sigma = 2 \cdot 4$).

Application of Normal Distribution

As it was said at the beginning, the normal law is very common in real life and thus has great practical application. Normal distribution is often used for practical measurements. Here are some examples of how it can be used.

Finding Probabilities

"Box of cereals" A packing machine is set to fill a cardboard box with a mean average of of cereal. Suppose the amounts per box form a normal distribution with a standard deviation equal to 0.04 ounce.

(a) What *percentage* of the boxes will end up with at least 1 pound of cereal?

(b) Ten percent of the boxes will contain more than what number of ounces?

This problem for instance has application in the industrial process of packing cereals and can be used for setting up the conveyor machine.

Solution:

Let X be the amount of cereals in the box (in ounces).

X is distributed normally: $X \sim N(16.1, 0.04^2)$

(a) We need to find the probability that X would be at least 16 ounces (1 pound=16 ounces), that is more or equal 16:

$$P(X \geqslant 16) = P(Z \geqslant \frac{16-\mu}{\sigma}) = P(Z \geqslant \frac{16-16.1}{0.04}) =$$
$$= P(Z \geqslant -2.5) = 1 - P(Z \geqslant -2.5) = 1 - 0.0062 = 0.9938$$

Notice that we used linear transformation that $z = \frac{X-\mu}{\sigma}$.

(b) Here we are given the probability and we need to find the critical raw score $0.1 = P(X < c)$ find c-?

$0.1 = P(Z < \frac{c-16.1}{0.04})$ and z = - 1.28 since 0.1 is to the left of exactly this value

$\frac{c-16.1}{0.04} = -1.28$

Converting it to the raw score gives: $c = \mu + z\sigma = 16.1 - 1.28 \cdot 0.04 = 16.049$

Finding Parameters Knowing Probabilities

Sometimes we are given the information about percentages of the distribution and one parameter. Given that it is normally distributed it is enough to calculate the unknown parameter, mean or standard deviation.

Example. Given a normal distribution with a mean of 25, what is its standard deviation, if 18% of the values are above 29?

Solution:

$$X \sim N(25, \sigma^2),\ \sigma = ?$$

We are given that $P(X \geqslant 29) = 0.18$

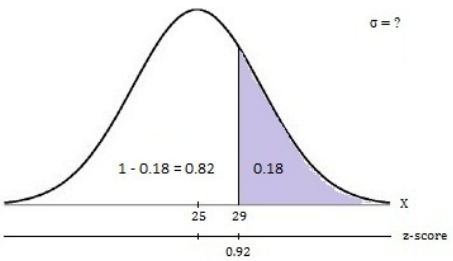
First find the z value corresponding to this area:

$P(Z \geqslant z) = 0.18$, then $P(Z \leqslant z) = 1 - 0.18 = 0.82$

Searching 0.82 in the table gives $z = 0.92$ since $P(Z \leqslant 0.92) = 0.82$.

Plugging into equation for z find required σ:

$Z = \frac{X-\mu}{\sigma}; 0.92 = \frac{29-25}{\sigma} \Rightarrow 29 - 25 = 0.92\sigma \Rightarrow \sigma = 4/0.92 = 4.35$



Analogously, knowing sigma we can find unknown μ.

Sum of Normal Random Variables

Theorem:
The Sum of *independent* Normal Random Variables is also a Normal Random Variable.

The crucial point is that the variables should be *independent*!

For example, consider three normal random variables A, B, C:
$A \sim N(\mu_a, \sigma_a^2)$
$B \sim N(\mu_b, \sigma_b^2)$
$C \sim N(\mu_c, \sigma_c^2)$,
where A, B and C are independent.

Let X be the average of those three variables: $X = \frac{A+B+C}{3}$.

Then X is *also* a *normal* random variable: $X \sim N(\mu_x, \sigma_x^2)$ where $\mu_x = \frac{a+b+c}{3}$
$\sigma_x^2 = Var\left(\frac{1}{3}\right)^2 (\sigma_a^2 + \sigma_b^2 + \sigma_c^2)$
Why we found μ_x and σ_x^2 like that? Recall the usual properties of expectation and variance and try to derive it by your own.

You must be able to reproduce even being drunk

- $F(x) = P(X \leqslant x_a) = \int_{-\infty}^{x_a} f(x)dx$
- $f(x) = F'(x)$
- the probability that a continuous random variable will take *a particular value* is zero
- $P(a \leqslant X \leqslant b) = \int_a^b f(x)dx = F(b) - F(a)$
- Any normal random variable could be transformed to the *standard normal* random variable, namely $Z = \frac{X-\mu}{\sigma}$
- $X \sim N(\mu, \sigma^2)$ – normal
- $Z \sim N(0, 1^2)$ – standard normal
- Normal curve is bell-shaped and symmetric around mean
- law of 1,2,3 sigmas with 68%, 95% and 99.7% correspondingly

Calculator BOX

- For calculating normal probabilities:
  `STAT` $\rightarrow$ `DIST` $\rightarrow$ `Norm` $\rightarrow$ `cd`
  Hint: there is no ∞ in your calculator, so insert just some very big number like 99999999.
- To find z value knowing probability:
  `STAT` $\rightarrow$ `DIST` $\rightarrow$ `Norm` $\rightarrow$ `Inverse`
  Choose Left tail if area is to the left of z $P(Z < c)$, right tail if area is to the right $P(Z > c)$

You can apply a standard strategy for solving this type of problems.

Full score strategy:

1. "Let it be" part
2. Write down how X is distributed
3. Reduce to z: e.g. $P(X < x) = P(Z < c)$
4. Provide the answer

Top secret information

- The expectation $E(X)$ of a continuous random variable X with density $f_X(x)$ using integral calculus is:

$$E(X) = \mu = \int_{-\infty}^{+\infty} x f_X(x) \mathrm{dx}$$

  This is similar though to the formula of discrete random variable expectation, where we multiply each value by its probability and sum those products. Here we also multiply each value by its probability (probability density) and take the integral which is analogue of sum.

- Formula for variance is also an analogue of a discrete case formula of variance

$$Var(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 f_X(x) \mathrm{dx}$$

Further recall more convenient formula that we used $Var(X) = E(X-\mu)^2 = E(X^2) - \mu^2$

How can we find $E(X^2)$ in continuous case?

Remind the property 5 of the expectation. If X is a random variable and there exist a continuous function $g(.)$, then $g(X)$ is also a random variable. Its expectation is: $E(g(X)) = \int_{-\infty}^{+\infty} g(x) f_X(x) dx$

We have $g(.) = X^2$ which is indeed continuous. Thus,

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f_X(x) dx$$

Plugging into formula for variance we get:

$$Var(X) \int_{-\infty}^{+\infty} x^2 f_X(x) dx - \mu^2$$

How did we get $Z = \frac{X-\mu}{\sigma}$?

Proof

We want to find a linear transformation $aX + b$ of normal variable X such that $aX + b = Z \sim N(0, 1^2)$ Recall from Chapter 2 that by properties of expectation and variance it is true that: $E(aX + b) = aE(X) + b$ and $Var(aX + b) = a^2 Var(X)$.

First, linear transformation of a normal variable will result in other normal variable. Let's find a and b such that $E(aX + b) = 0$ and $Var(aX + b) = 1$.

$$aE(X) + b = 0$$

$$a^2 Var(X) = 1$$

$$a = \sqrt{\frac{1}{Var(X)}} = \frac{1}{\sigma}$$

$$b = -aE(X) = -\frac{1}{\sigma}E(X) = \frac{1}{\sigma}\mu$$

Thus, for $a = \frac{1}{\sigma}$ and $b = -\frac{\mu}{\sigma}$ we will get a random variable with mean 0 and standard deviation 1!

Plug it into expression $aX + b$:

$aX + b = \frac{1}{\sigma}X - \frac{1}{\sigma}\mu = \frac{X-\mu}{\sigma} = Z$

Thus, $Z = \frac{X-\mu}{\sigma} \sim N(0, 1^2)$

So, subtracting from any random variable its mean and dividing by its standard deviation yields a random variable with mean 0 and standard deviation 1.

Sample AP problems with solutions

Problem 1. AP 2017 №3
A grocery store purchases melons from two distributors, J and K. Distributor J provides melons from organic farms. The distribution of the diameters of the melons from Distributor J is approximately normal with mean 133 millimeters (mm) and standard deviation 5 mm.

(a) For a melon selected at random from Distributor J, what is the probability that the melon will have a diameter greater than 137 mm?

Solution

Let X denote the diameter of a randomly chosen melon from Distributor J. X has an approximately normal distribution with mean 133 mm and standard deviation 5 mm.

The z-score for a diameter of 137 mm is $z = \frac{137-133}{5} = \frac{4}{5} = 0.8$.

Therefore, $P(X > 137) = P(Z > 0.8) = 1 - 0.7881 = 0.2119$.

Problem 2. AP 2014 №3
Schools in a certain state receive funding based on the number of students who attend the school. To determine the number of students who attend a school, one school day is selected at random and the number of students in attendance that day is counted and used for funding purposes. The daily number of absences at High School A in the state is approximately normally distributed with mean of 120 students and standard deviation of 10.5 students.

(a) If more than 140 students are absent on the day the attendance count is taken for funding purposes, the school will lose some of its state funding in the subsequent year. Approximately what is the probability that High School A will lose some state funding?

The primary goal of this question was to assess a student's ability to perform a probability calculation from a normal distribution

Solution

(a) Let X be the daily number of absences.
$X \sim N(120, 10.5^2)$
High School will lose some state funding if number of daily absences would be more than 140. Thus we need to find:

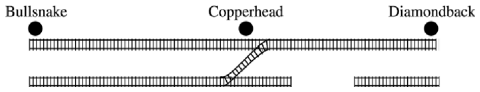
$$P(X \geqslant 140) = P(Z \geqslant 140 - 120/10.5) = P(Z \geqslant 1.9) =$$
$$= 1 - P(Z \leqslant 1.9) = 1 - 0.9713 = \mathbf{0.0287}$$

Answer: Approximate probability that the school will lose some state funding

is 0.0287.

Problem 2. AP 2008 Form B №5

Flooding has washed out one of the tracks of the Snake Gulch Railroad. The railroad has two parallel tracks from Bullsnake to Copperhead, but only one usable track from Copperhead to Diamondback, as shown in the figure below. Having only one usable track disrupts the usual schedule. Until it is repaired, the washed-out track will remain unusable. If the train leaving Bullsnake arrives at Copperhead first, it has to wait until the train leaving Diamondback arrives at Copperhead.



Every day at noon a train leaves Bullsnake heading for Diamondback and another leaves Diamondback heading for Bullsnake.

Assume that the length of time, X, it takes the train leaving Bullsnake to get to Copperhead is normally distributed with a mean of 170 minutes and a standard deviation of 20 minutes.

Assume that the length of time, Y, it takes the train leaving Diamondback to get to Copperhead is normally distributed with a mean of 200 minutes and a standard deviation of 10 minutes.

These two travel times are independent.

(a) What is the distribution of $Y - X$?

(b) Over the long run, what proportion of the days will the train from Bullsnake have to wait at Copperhead for the train from Diamondback to arrive?

(c) How long should the Snake Gulch Railroad delay the departure of the train from Bullsnake so that the probability that it has to wait is only 0.01?

Intent of Question

The primary goals of this question were to assess a student's ability to (1) describe the distribution of the difference of two normal random variables and (2) use this distribution to find a probability and to find a value given its location in the distribution.

Solution:

(a) X is normally distributed with $\mu = 170$ and $\sigma = 20$, and Y is normally distributed with $\mu = 200$ and $\sigma = 10$:
$X \sim N(170, 20^2)$
$Y \sim N(200, 10^2)$

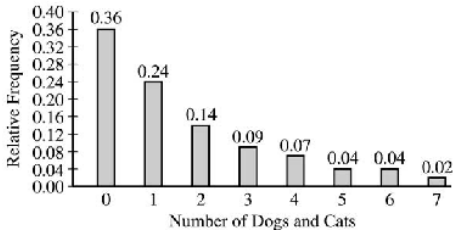
By the theorem, the sum of independent normal random variables is also a normal random variable, thus:
$Y - X \sim N(\mu, \sigma^2)$
$\mu_{Y-X} = E(X - Y) = 200 - 170 = 30$
$\sigma_{Y-X} = \sqrt{10^2 + 20^2} = \sqrt{500} = 22.36$

(b) The train from Bullsnake have to wait for the train from Diamondback when $Y > X$:
$P(Y > X) = P(Y - X > 0) = P(Z > \frac{0-30}{22.36}) = P(Z > -1.34) = 0.9099$

The proportion of days that the train from Bullsnake will have to wait is about 0.91.

(c) Let D denote the delay that will be needed for the train leaving Bullsnake. D is a constant. Denote $X' = X + D$
$X' \sim N(\mu_{X'}, \sigma^2_{X'})$
$\mu_{X'} = 170 + D$
$\sigma_{X'} = \sigma_X = 20$
Thus, the difference $Y - X' \sim N(\mu_{Y-X'}, \sigma_{Y-X'})$
$\mu_{Y-X'} = 200 - (170 + D) = 200 - 170 - D = 30 - D$
$\sigma_{Y-X'} = \sigma_{Y-X} = 22.36$
The travel time plus delay X' for the Bullsnake train must be *less than* the travel time Y for the Diamondback train with probability 0.01: $P(Y > X') = P(Y - X' > 0) = 0.01$
$P(Z > z) = 0.01$
checking for z-score gives $z = 2.33$ since $P(Z > 2.33) = 0.01$
$2.33 = \frac{0-(30-D)}{22.36} \Rightarrow D - 30 = 52.0988 \Rightarrow D = 82.0988$ Answer: The train from Bullsnake should be delayed by 82.099 minutes.

Problem 3. AP 2007 Form B №2
The graph below displays the relative frequency distribution for X, the total number of dogs and cats owned per household, for the households in a large suburban area. For instance, 14 percent of the households own 2 of these pets.



(a) According to a local law, each household in this area is prohibited from owning more than 3 of these pets. If a household in this area is selected at random, what is the probability that the selected household will be in violation of this law? Show your work.

Solution:
$P(X > 3) = 0.07 + 0.04 + 0.04 + 0.02 = 0.17$

Practice AP problems

Problem 1. AP 2013 №3
Each full carton of Grade A eggs consists of 1 randomly selected empty cardboard container and 12 randomly selected eggs. The weights of such full cartons are approximately normally distributed with a mean of 840 grams and a standard deviation of 7.9 grams.

(a) What is the probability that a randomly selected full carton of Grade A eggs will weigh more than 850 grams?

(b) The weights of the empty cardboard containers have a mean of 20 grams and a standard deviation of 1.7 grams. It is reasonable to assume independence between the weights of the empty cardboard containers and the weights of the eggs. It is also reasonable to assume independence among the weights of the 12 eggs that are randomly selected for a full carton. Let the random variable X be the weight of a single randomly selected Grade A egg.

i) What is the mean of X?

ii) What is the standard deviation of X?

Problem 2. AP 2006 №3 The depth from the surface of Earth to a refracing layer beneath the surface can be estimated using methods developed by seismologists. One method is based on the time required for vibrations to travel from a distant explosion to a receiving point. The depth measurement M is the sum of the true depth D and the random measurement E. That is, $M = D + E$. The measurement error E is assumed to be normally distributed with mean 0 feet and standard deviation 1.5 feet.

(a) If the true depth at a certain point is 2 feet, what is the probability that the depth measurement will be negative?

Answers to the practice problems:

1. (a) 0.1020 (b) 68.33 gr, 2.23 gr
2. 0.0918