Technische Universiteit Delft
Faculteit Elektrotechniek, Wiskunde en Informatica
Delft Institute of Applied Mathematics

Over verschillende karakteriseringen van een normaal verdeling

(Engelse titel: On different characterizations of a normal distribution)

Verslag ten behoeve van het
Delft Institute of Applied Mathematics
als onderdeel ter verkrijging

van de graad van

BACHELOR OF SCIENCE
in
TECHNISCHE WISKUNDE

door

HIDDE VAN WIECHEN

Delft, Nederland
Augustus 2018

BSc verslag TECHNISCHE WISKUNDE

"Over verschillende karakteriseringen van een normaal verdeling"

(Engelse titel: "On different characterizations of a normal distribution")

HIDDE VAN WIECHEN

Technische Universiteit Delft

Begeleider

Prof.Dr.ir. M.C. Veraar

Overige commissieleden

Dr. W.M. Ruszel

Dr. J.A.M. de Groot

Augustus, 2018

Delft

Abstract

The normal distribution is a very important distribution in probability theory and statistics and has a lot of unique properties and characterizations. In this report we look at the proof of two of these characterizations and create counterparts of a normal distribution on abstract spaces, such as vector spaces and groups, which we shall call Gaussians. When we look at $\mathbb{R}^d$, all these Gaussians coincide, along with a Gaussian vector in the normal sense, called multivariate normal. Furthermore, for one Gaussian we prove that it has exponential integrability properties.

Contents

Introduction

When people think about the normal distribution a lot of interesting properties come to mind; how its density looks like a bell, the easy way to determine the 95% interval using the standard deviation, the central limit theorem, etc. There are however some lesser known properties that can even define a normal distribution on $\mathbb{R}$. These so-called characterizations are at first glance not obviously linked to the normal distribution but apparently have a very strong connection with it.

Once one has looked at these characterizations, a question that immediately emerges from this is what happens when we make definitions of these characterizations and look at them on different spaces than just $\mathbb{R}$. This question is the central point of this report and will be examined.

This research all started with an article by K. Oleszkiewicz [9] and a book by H. Bauer [2], which both look at different characterizations that we will be discussing in this report. W. Bryc has also examined these characterizations and wrote in his lecture notes about the definitions that can be derived from them [5]. Since his notes go deeper than the material in this report, they can be seen as an extension to be read after. A summary of each chapter follows below.

Chapter one gives the basic knowledge of certain subjects that are needed in this report. Most of the information in this chapter is being taught in many probability courses, and otherwise we will take a deeper look at it in here. A basic understanding of probability and analysis is needed for this chapter.

In chapter two we prove the characterizations that are central in this report. We will also have to look at other theorems, propositions and lemmas needed for these proofs, including Cramér's theorem, which is also an interesting property of the normal distribution.

Chapter three is all about the different types of a Gaussian that we define, using the characterizations from chapter two. We will see on what kind of abstract spaces we can define these Gaussians and when the definitions coincide with one another. Lastly, we will prove the integrability of one of these Gaussians.

Chapter 1

Prior Knowledge

Before we take a look at the characterizations, we will need some prior knowledge on certain probability tools and the normal distribution. A lot of the propositions and theorems in this section are proved in many probability courses, such as *Advanced probability (TW3560)* [10] given at Delft University of Technology.

1.1 Expectation and moments

Definition 1.1.
Let X be a random variable on sample space Ω with probability measure P, then we write the expectation as

$$\mathbb{E}(X) = \int_{\Omega} X dP.$$

This definition is very abstract, however we can write this in ways that are more useful. For instance with the help of the cumulative distribution function $F(x) = \mathbb{P}(X \leq x)$,

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x dF(x),$$

or even better, if X has density $f(x)$, then

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx.$$

Some well known properties of the expectation are listed below

Proposition 1.1. (Basic properties of expectation)
Let X, Y be random variables, then

1. *for $a, b \in \mathbb{R}$, $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$.*
2. *if $Y \leq X$ then $\mathbb{E}(Y) \leq \mathbb{E}(X)$.*
3. *if X and Y are independent, then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.*
4. *$|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$.*

Another great property is the inequality known as Jensen's inequality.

Theorem 1.2. **(Jensen's inequality)**
Let X be a random variable and $g : \mathbb{R} \to \mathbb{R}$ a convex function such that X and $g(X)$ are integrable, then

$$g\left(\mathbb{E}(X)\right) \leq \mathbb{E}(g(X)).$$

The 4^{th} statement of Proposition 1.1 is a special case of this inequality. Another very common example where Jensen's inequality can be used is to show that $\mathbb{E}(X)^2 \leq \mathbb{E}(X^2)$, and therefore it is obvious that $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \geq 0$.

The following proposition will be very useful later.

Proposition 1.3.
If X is a non-negative random variable and $h : \mathbb{R} \to \mathbb{R}$ a function with $h'(x) \geq 0$ for $x \geq 0$, such that $\mathbb{E}(h(X)) < \infty$, then

$$\mathbb{E}(h(X)) = \int_0^\infty \mathbb{P}(X > x)h'(x)dx + h(0).$$

Proof. We can write $\mathbb{P}(X > x) = \mathbb{E}\left(\mathbb{1}_{X>x}\right)$, thus

$$\int_0^\infty \mathbb{P}(X > x)h'(x)dx = \int_0^\infty \mathbb{E}\left(\mathbb{1}_{X>x}\right)h'(x)dx = \int_0^\infty \int_\Omega \mathbb{1}_{X>x}h'(x)dPdx.$$

Since $\mathbb{1}_{X>x}h'(x)$ is non-negative for $x \geq 0$ and measurable, we are able to use Fubini's theorem [1, chapter 4] for changing the order of integration.

$$\begin{aligned}
\int_\Omega \int_0^\infty \mathbb{1}_{X>x}h'(x)dxdP &= \int_\Omega \int_0^X h'(x)dxdP \\
&= \int_\Omega (h(X) - h(0))dP \\
&= \mathbb{E}(h(X)) - h(0),
\end{aligned}$$

therefore

$$\int_0^\infty \mathbb{P}(X > x)h'(x)dx = \mathbb{E}(h(X)) - h(0).$$

Sending $h(0)$ over to the other side gives us the needed result. □

Now that we have talked about expectations we can also talk about moments. The moment of order $n = 0, 1, 2, \ldots$ is defined by $\mathbb{E}(X^n)$ and the absolute moment of order $n \geq 0$ by $\mathbb{E}(|X|^n)$. If the absolute moment of order n exists, i.e. $\mathbb{E}(|X|^n) < \infty$, then the moment of order k with $0 \leq k \leq n$ also exists. However, there is a way to find out if all the moments of a random variable exist, using the moment-generating function given by

$$M_X(t) = \mathbb{E}\left(e^{tX}\right).$$

If the moment-generating function of a random variable exists, for at least an area around zero, then so do all its moments. This has to do with the fact that if $\mathbb{E}\left(e^{tX}\right) < \infty$ for $|t| < \varepsilon$ then $\mathbb{E}\left(e^{t|X|}\right) < \infty$ for the same t. To see this, we split the integral in two parts.

$$\mathbb{E}\left(e^{t|X|}\right) = \int_{\Omega} e^{t|X|} dP = \int_{X\geq 0} e^{tX} dP + \int_{X<0} e^{-tX} dP.$$

Since $\mathbb{E}\left(e^{tX}\right) = \int_{\Omega} e^{tX} dP < \infty$ for $|t| < \varepsilon$, and $e^x > 0$ for all $x \in \mathbb{R}$, the two smaller integrals above have to be finite as well. Combining this with the fact that $t^n \leq n! e^t$ for all $t \geq 0$ (this follows immediately after substituting $e^t = \sum_{k=0}^{\infty} \frac{t^k}{k!}$), we see that

$$t^n \mathbb{E}(|X|^n) \leq n! \mathbb{E}\left(e^{t|X|}\right) < \infty,$$

therefore, X has finite (absolute) moments. The moment-generating function also gives a new way to calculate these moments.

$$M_X(t) = \mathbb{E}\left(\sum_{n=0}^{\infty} \frac{t^n X^n}{n!}\right) = \int_{\Omega} \sum_{n=0}^{\infty} \frac{t^n X^n}{n!}.$$

Now suppose there is an $\varepsilon > 0$ such that if $M_X(t)$ is finite for $|t| < \varepsilon$, then Fubini's theorem tells us we can interchange the integral and infinite sum for these t , so

$$M_X(t) = \sum_{n=0}^{\infty} \frac{t^n \int_{\Omega} X^n}{n!} = \sum_{n=0}^{\infty} \frac{t^n \mathbb{E}(X^n)}{n!}.$$

If we now would take the k'th derivative of $M_X(t)$ and substitute $t = 0$, then this series tells us that this would be equal to $\mathbb{E}(X^k)$.

1.2 Characteristic functions

The characteristic function of a random variable is defined as follows.

Definition 1.2.
The characteristic function of a real-valued random variable X is given by

$$\varphi_X(t) = \mathbb{E}\left(e^{itX}\right).$$

For a continuous random variable with density $f(x)$ this can also be seen as the Fourier-transform of this density. As you can see it is almost the same as the moment-generating function, however the beauty of the characteristic function of a random variable is that it always exists, since $\left|e^{itX}\right| = 1$ for all t thus it is integrable, and comes with some very useful properties that we will be discussing in this section. The most important one is the uniqueness of the characteristic function.

Theorem 1.4. (Uniqueness theorem)
Let X and Y be two random variables, then $\varphi_X = \varphi_Y$ if and only if $X \overset{d}{=} Y$.

This uniqueness theorem tells us that the distribution of a random variable is determined uniquely by its characteristic function. With this we will be able to make a more useful definition of a normal random variable, as seen in the next section.

Some other, basic properties of a characteristic function that we will need are the following.

Proposition 1.5. **(Basic properties of characteristic functions)**
Let X be a random variable with characteristic function $\varphi_X(t)$ and let $t \in \mathbb{R}$, then

1. $\overline{\varphi_X(t)} = \varphi_X(-t)$.
2. $|\varphi_X(t)| \leq \varphi_X(0) = 1$.
3. $\varphi_X(t)$ *is uniformly continuous.*

Much like the moment generating function, we can find a connection with the moments of X. If $\mathbb{E}(|X|^k) < \infty$ then we can find that the k'th derivative of φ_X exists and is equal to

$$\varphi_X^{(k)}(t) = \frac{d^k}{dt^k}\mathbb{E}\left(e^{itX}\right) = \mathbb{E}\left((iX)^k e^{itX}\right).$$

So we are able to switch the derivative and expectation. An interesting case is when we substitute $t = 0$

$$\varphi_X^{(k)}(0) = i^k \mathbb{E}(X^k). \tag{1.2.1}$$

This trick also works the other way around.

Proposition 1.6.
If for even k the k'th derivative of φ_X exists, then $\mathbb{E}(X^k) < \infty$ and (1.2.1) still applies

Harald Cramér has given a proof of this [6, chapter 10]. However, this only works for even k. Antoni Zygmund has given an example of a characteristic function which has a derivative, but the expected value of the random variable corresponding to it does not exist [12].

Another great thing about the characteristic functions is what happens if independence is given. Suppose $X_1, \ldots, X_n$ are independent random variables and $a_1, \ldots, a_n$ some constants, then

$$\begin{aligned}\varphi_{\sum_{k=1}^n a_k X_k}(t) &= \mathbb{E}\left(e^{it\sum_{k=1}^n a_k X_k}\right) \\ &= \prod_{k=1}^n \mathbb{E}\left(e^{ita_k X_k}\right) \qquad (1.2.2) \\ &= \prod_{k=1}^n \varphi_{X_k}(a_k t).\end{aligned}$$

So if we are dealing with independent random variables, the characteristic function of the sum is equal to the product of the characteristic functions. This is a lot easier than, for example, density functions where we would have to use convolutions to determine the new density.

The reverse however is not true; if $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$, then X and Y don't necessarily have to be independent. To determine whether X and Y are independent, we could have a look at the characteristic functions of *random vectors*. If we would have multiple random variables $X_1, X_2, ..., X_d$ and put them all in a vector $\mathbf{X}$ we would have what is called a random vector in $\mathbb{R}^d$. For $\mathbf{X}$ a d-dimensional random vector, the characteristic function is defined as

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}\left(e^{i\mathbf{t}^T\mathbf{X}}\right)$$

with $\mathbf{t} \in \mathbb{R}^d$. This way we can give the following sufficient condition for independence.

Proposition 1.7.
The $\mathbb{R}^d$ valued random vectors $\mathbf{X}$,$\mathbf{Y}$ are independent if and only if the equality

$$\varphi_{(\mathbf{X},\mathbf{Y})}(\mathbf{s},\mathbf{t}) = \varphi_{\mathbf{X}}(\mathbf{s})\varphi_{\mathbf{Y}}(\mathbf{t}). \tag{1.2.3}$$

holds for all $\mathbf{s}, \mathbf{t} \in \mathbb{R}^d$.

The proof of this proposition is almost trivial. If we would assume that (1.2.3) holds then, due to the uniqueness theorem, the joint distribution function is equal to the product of the marginal distribution functions, and therefore $\mathbf{X}$ and $\mathbf{Y}$ are independent. Furthermore, if $\mathbf{X}$ and $\mathbf{Y}$ are independent, then we can show that (1.2.3) holds, much like (1.2.2).

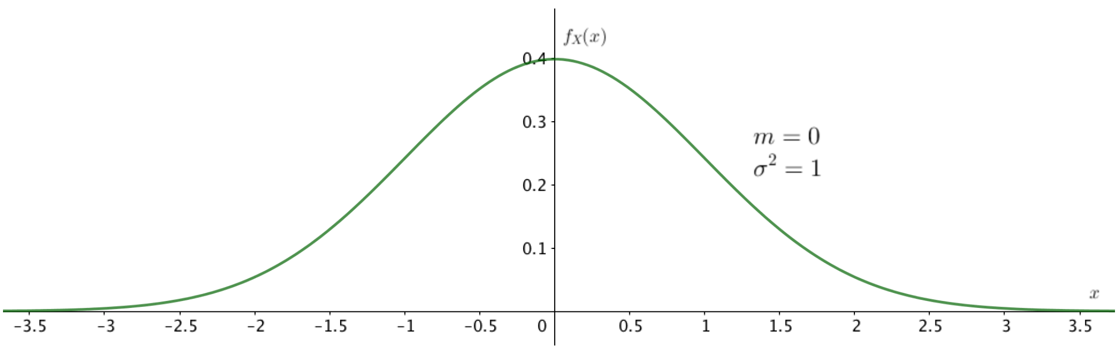
1.3 Normal distribution

Suppose X is a normally distributed random variable with mean m and variance σ^2, then we write this as $X \sim N(m, \sigma^2)$. Usually an X like this is defined by its density

$$f_X(x) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

This function gives us the well known, bell-shaped curve associated with the normal distribution as seen in the figure below.

Figure 1.1: Density function of a standard normal variable.



However we will be working a lot with independent sums of normal random variables, and if we would then work with the density functions, we would have to calculate difficult convolutions which would take a lot of time. As we have seen in the previous section, these sums are easier to analyse with the help of the characteristic functions. Fortunately, due to the uniqueness theorem, we are able to define a normal random variable by its characteristic function.

So we need the characteristic function of a normal distribution. It has been proven that the characteristic function of a standard normal random variable $Z \sim N(0,1)$ is equal to $\varphi_Z(t) = e^{-\frac{t^2}{2}}$ (see [10, chapter 9]). From this we can calculate the characteristic function of any normally distributed random variable.

Suppose X has a normal distribution with mean m and variance σ^2, then we can write $X = \sigma Z + m$, Therefore

$$\varphi_X(t) = \mathbb{E}\left(e^{i\sigma tZ+tm}\right) = e^{itm}\mathbb{E}\left(e^{i\sigma tZ}\right) = e^{itm}\varphi_Z(\sigma t).$$

Since we know that characteristic function of Z, we end up with

$$\varphi_X(t) = e^{itm-\frac{1}{2}\sigma^2t^2}.$$

With this characteristic function in our mind, we can give the following definition of a normal distribution.

Definition 1.3.
A real valued random variable X has the normal distribution $N(m,\sigma^2)$ if its characteristic function has the form

$$\varphi_X(t) = e^{itm-\frac{1}{2}\sigma^2t^2}. \tag{1.3.1}$$

This definition gives us the following proposition.

Proposition 1.8.
A characteristic function which can be expressed in the form

$$\varphi_X(t) = e^{at^2+bt+c}$$

for some $a, b, c \in \mathbb{C}$, corresponds to a normal distribution.

Proof. Suppose that the characteristic function of a random variable X is as in the proposition above. Since φ_X is twice differentiable, Proposition 1.6 tells us that X has an expectation and variance, which we will denote as m and σ^2 respectively. First of all, Proposition 1.5 tells us that $\varphi_X(0) = 1$, so we can conclude that $c = 0$. Furthermore, the first and second derivative of φ_X are equal to

$$\begin{aligned}
\varphi_X'(t) &= (2at+b)e^{at^2+bt}, \\
\varphi_X''(t) &= (2at+b)^2e^{at^2+bt} + 2ae^{at^2+bt}.
\end{aligned}$$

Substituting $t = 0$ and using (1.2.1), we get

$$\varphi'_X(0) = i\mathbb{E}(X) = b,$$
$$\varphi''_X(0) = -\mathbb{E}(X^2) = b^2 + 2a.$$

So $b = im$ and thus

$$a = \frac{1}{2}\left(m^2 - \mathbb{E}(X^2)\right) = -\frac{1}{2}\sigma^2.$$

In conclusion,

$$\varphi_X(t) = e^{-\frac{1}{2}\sigma^2 t^2 + itm},$$

therefore X has a normal distribution. □

1.4 Multivariate normal

If we would look at random vectors we also have a normal distribution, commonly known as *multivariate normal*.

Definition 1.4.
An $\mathbb{R}^d$-valued random vector $\mathbf{X}$ is multivariate normal if for every $\mathbf{t} \in \mathbb{R}^d$ the real valued random variable $\mathbf{t}^T\mathbf{X}$ is normal.

The favorable thing that this definition gives us is that we make a new, so-called *univariate* normal random variable, namely $\mathbf{t}^T\mathbf{X}$ on $\mathbb{R}$, about which we know a lot. Also, when we take $\mathbf{t}$ such that $t_i = 1$ and $t_j = 0$ for $j \neq i$, then it is easy to see that all the X_i have to be univariate normal.

Since $\mathbf{t}^T\mathbf{X}$ is a normal random variable it has a mean $m_\mathbf{t}$ and variance $\sigma^2_\mathbf{t}$. Due to the linearity of the expecation, we can deduce that $m_\mathbf{t} = \mathbf{t}^T\mathbf{m}$, with $\mathbf{m} = \mathbb{E}(\mathbf{X})$, the vector where the i'th component equals $\mathbb{E}(X_i)$ for $1 \leq i \leq d$. The variance however isn't that easy. We will first look at the situation where $d = 2$. If we would have two normally distributed random variables $X_1 \sim N(m_1, \sigma_1^2)$ and $X_2 \sim N(m_2, \sigma_2^2)$, then the variance of $t_1X_1 + t_2X_2$ would be equal to

$$\sigma^2_\mathbf{t} = t_1^2\sigma_2^2 + t_2^2\sigma_2^2 + 2t_1t_2\mathrm{Cov}(X_1, X_2),$$

after using the properties of the variance and covariance. We can simplify this a bit by writing it in matrix form

$$\sigma^2_\mathbf{t} = \mathbf{t}^T\Sigma\mathbf{t}, \tag{1.4.1}$$

with

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \mathrm{Cov}(X_1, X_2) \\ \mathrm{Cov}(X_2, X_1) & \sigma_2^2 \end{bmatrix}.$$

This Σ is commonly known as the *covariance matrix*, where the element on the i'th row of the j'th collumn is $\mathrm{Cov}(X_i, X_j)$ (notice that when $i = j$ this is equal to $\mathrm{Cov}(X_i, X_i) = \sigma_i^2$).

If we would look at the variance for general d we would again be able to write it as (1.4.1), but with a $d \times d$ covariance matrix Σ. With this, we are able to determine the characteristic function of $\mathbf{t}^T\mathbf{X}$.

$$\varphi_{\mathbf{t}^T\mathbf{X}}(s) = \mathbb{E}\left(e^{i\mathbf{t}^T\mathbf{X}s}\right) = \exp\left(-\frac{1}{2}\mathbf{t}^T\Sigma\mathbf{t}s^2 + i\mathbf{t}^T\mathbf{m}s\right).$$

Taking $s = 1$ and turning $\mathbf{t}$ into a variable gives

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}\left(e^{i\mathbf{t}^T\mathbf{X}}\right) = \exp\left(-\frac{1}{2}\mathbf{t}^T\Sigma\mathbf{t} + i\mathbf{t}^T\mathbf{m}\right), \tag{1.4.2}$$

which is the characteristic function of our random vector $\mathbf{X}$. Therefore we now have the characteristic function corresponding to the multivariate normal distribution. From now on we will write $\mathbf{X} \sim \mathcal{N}(\mathbf{m}, \Sigma)$ to say that $\mathbf{X}$ is multivariate normal with mean $\mathbf{m}$ and covariance matrix Σ.

The last proposition has to do with independent multivariate normal random variables.

Proposition 1.9.
If $\mathbf{X}$ and $\mathbf{Y}$ are independent multivariate normal random variables, then $\mathbf{X} + \mathbf{Y}$ is still multivariate normal.

Proof. Assume $\mathbf{X} \sim \mathcal{N}(\mathbf{m}_1, \Sigma_1)$ and $\mathbf{X} \sim \mathcal{N}(\mathbf{m}_2, \Sigma_2)$ are independent, then we know that

$$\begin{aligned}
\varphi_{\mathbf{X}+\mathbf{Y}}(\mathbf{t}) &= \varphi_{\mathbf{X}}(\mathbf{t})\varphi_{\mathbf{Y}}(\mathbf{t}) \\
&= \exp\left(-\frac{1}{2}\mathbf{t}^T\Sigma_1\mathbf{t} + i\mathbf{t}^T\mathbf{m}_1\right) \cdot \exp\left(-\frac{1}{2}\mathbf{t}^T\Sigma_2\mathbf{t} + i\mathbf{t}^T\mathbf{m}_2\right) \\
&= \exp\left(-\frac{1}{2}\mathbf{t}^T(\Sigma_1 + \Sigma_2)\mathbf{t} + i\mathbf{t}^T(\mathbf{m}_1 + \mathbf{m}_2)\right)
\end{aligned}$$

Therefore we can see that $\mathbf{X} + \mathbf{Y} \sim \mathcal{N}(\mathbf{m}_1 + \mathbf{m}_2, \Sigma_1 + \Sigma_2)$, hence it is multivariate normal. □

Chapter 2

Characterizations of a normal distribution on $\mathbb{R}$

In this chapter we will take a deeper look at the normal distribution by proving certain characterizations, which we will do in section 2.1 and 2.3. Section 2.2 is all about Cramér's theorem, which will be needed for proving the second characterization.

2.1 First characterization

In this section we will look at the first characterization. Although it has been proven in different ways, one of which can be read in [5, section 5.1], the proof in this section is derived from the one by H. Bauer [2, section 24] since it mainly focusses on characteristic functions, about which we have learned a lot in the previous chapter.

Theorem 2.1. **(First characterization)**
If X and Y are two real valued independent random variables, the following statements are equivalent:

1. *X and Y are normally distributed with equal variance.*
2. *The random variables $X + Y$ and $X - Y$ are independent.*

Since it is an equivalence, it will be necessary to prove both directions. Let's start with the easy one.

Proof of 1 ⇒ 2.
Assume that $X \sim N(\alpha, \sigma^2)$ and $Y \sim N(\beta, \sigma^2)$ are independent. First of all,

$$X + Y \sim N(\alpha + \beta, 2\sigma^2)$$
$$X - Y \sim N(\alpha - \beta, 2\sigma^2)$$

To prove the independence of $X + Y$ and $X - Y$, it is possible to use proposition 1.7. That way, all we have to show is that for all $s, t \in \mathbb{R}$,

$$\varphi_{(X+Y,X-Y)}(s, t) = \varphi_{X+Y}(s) \cdot \varphi_{X-Y}(t).$$

First of all,

$$\varphi_{(X+Y,X-Y)}(s,t) = \mathbb{E}\left[e^{i(s(X+Y)+t(X-Y))}\right] = \mathbb{E}\left[e^{i(s+t)X}e^{i(s-t)Y}\right],$$

and since X and Y are independent,

$$\mathbb{E}\left[e^{i(s+t)X}e^{i(s-t)Y}\right] = \mathbb{E}\left[e^{i(s+t)X}\right]\mathbb{E}\left[e^{i(s-t)Y}\right] = \varphi_X(s+t)\cdot\varphi_Y(s-t).$$

Luckily, the distributions of X and Y are given, and therefore also their characteristic functions, so

$$\begin{aligned}\varphi_X(s+t)\cdot\varphi_Y(s-t) &= e^{i\alpha(s+t)-\frac{\sigma^2(s+t)^2}{2}}\cdot e^{i\beta(s-t)-\frac{\sigma^2(s-t)^2}{2}}\\ &= e^{i(\alpha+\beta)s-\sigma^2s^2}\cdot e^{i(\alpha-\beta)t-\sigma^2t^2}\\ &= \varphi_{X+Y}(s)\cdot\varphi_{X-Y}(t),\end{aligned}$$

which concludes the proof. □

The other direction is going to require some more effort. We are first going to prove the following lemma.

Lemma 2.2.
Let $\chi : \mathbb{R} \to \mathbb{T}$, with $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$, be a continuous function such that $\chi(a+b) = \chi(a)\chi(b)$ for all $a, b \in \mathbb{R}$. Then

$$\chi(x) = e^{i\alpha x}$$

for some $\alpha \in \mathbb{R}$

Proof. For all $a \in \mathbb{R}$

$$\chi(a) = \chi(a+0) = \chi(a)\chi(0),$$

therefore $\chi(0) = 1$. Take $c \in \mathbb{R}$ such that

$$b := \int_0^c \chi(t)dt \neq 0,$$

then for all $x \in \mathbb{R}$,

$$b\cdot\chi(x) = \int_0^c \chi(t)\chi(x)dt = \int_0^c \chi(t+x)dt = \int_x^{x+c}\chi(t)dt.$$

Taking the derivative to x on both sides gives

$$b\cdot\chi'(x) = \frac{d}{dx}\left(\int_x^{x+c}\chi(t)dt\right).$$

Define $F(y) := \int_{x_0}^{y} \chi(t)dt$ with $x_0, y \in [x, x+c]$, then it follows that

$$\begin{aligned} b \cdot \chi'(x) &= \frac{d}{dx}\left(\int_{x_0}^{x+c} \chi(t)dt + \int_{x}^{x_0} \chi(t)dt\right) \\ &= \frac{d}{dx}\left(F(x+c) - F(x)\right) \\ &= \chi(x+c) - \chi(x) \\ &= \chi(x)\chi(c) - \chi(x) \\ &= \Big(\chi(c) - 1\Big)\chi(x), \end{aligned}$$

and so

$$\chi'(x) = \frac{\Big(\chi(c) - 1\Big)}{b}\chi(x) = C\chi(x)$$

for some $C \in \mathbb{C}$. The differential equation $\chi'(x) = C\chi(x)$ is a standard one which has the general solution

$$\chi(x) = Ae^{Cx},$$

and since $\chi(0) = 1$,

$$1 = \chi(0) = Ae^0 = A.$$

Therefore the function equals $\chi(x) = e^{Cx}$. Considering $\chi : \mathbb{R} \to \mathbb{T}$ it follows that $e^{Cx} \in \mathbb{T}$ for all $x \in \mathbb{R}$, so $\left|e^{Cx}\right| = 1$. C is a complex number so it can be written as $C = i\alpha + \beta$, with $\alpha, \beta \in \mathbb{R}$, which means that for all $x \in \mathbb{R}$,

$$\left|e^{Cx}\right| = \left|e^{i\alpha x + \beta x}\right| = \left|e^{i\alpha x}\right| \cdot \left|e^{\beta x}\right| = \left|e^{\beta x}\right| = 1.$$

This only holds for $\beta = 0$, therefore

$$\chi(x) = e^{i\alpha x},$$

which proves this lemma. $\square$

Continued proof of theorem 2.1.1, 2 ⇒ 1.
This proof will be given in three distinct parts followed up by the conclusion, in which we will show that X and Y are normal distributions according to Definition 1.3. The goal of the first part is to show that φ_X and φ_Y are never zero, which will be needed for the second and third part.

Part 1
Define $U := \frac{X+Y}{2}$ and $V := \frac{X-Y}{2}$. Then the equalities $X = U + V$ and $Y = U - V$ hold. Because of the independence of X and Y,

$$\varphi_X(x)\varphi_Y(y) = \mathbb{E}\left(e^{ixX}e^{iyY}\right) = \mathbb{E}\left(e^{ix(U+V)}e^{iy(U-V)}\right) = \mathbb{E}\left(e^{i(x+y)U}e^{i(x-y)V}\right).$$

The assumption tells us that U and V are also independent, therefore

$$\begin{aligned}\varphi_X(x)\varphi_Y(y) &= \mathbb{E}\left(e^{i(x+y)U}\right)\mathbb{E}\left(e^{i(x-y)V}\right)\\ &= \mathbb{E}\left(e^{i\frac{x+y}{2}X}e^{i\frac{x+y}{2}Y}\right)\mathbb{E}\left(e^{i\frac{x-y}{2}X}e^{-i\frac{x-y}{2}Y}\right)\\ &= \varphi_X\left(\frac{x+y}{2}\right)\varphi_Y\left(\frac{x+y}{2}\right)\varphi_X\left(\frac{x-y}{2}\right)\overline{\varphi_Y\left(\frac{x-y}{2}\right)}.\end{aligned} \tag{2.1.1}$$

Fill in $y = 0$ and you are left with

$$\varphi_X(x) = \varphi_X\left(\frac{x}{2}\right)^2\left|\varphi_Y\left(\frac{x}{2}\right)\right|^2,$$

which gives

$$|\varphi_X(x)| = \left|\varphi_X\left(\frac{x}{2}\right)\right|^2\left|\varphi_Y\left(\frac{x}{2}\right)\right|^2.$$

Repeat this process for φ_Y and you will find

$$|\varphi_Y(x)| = \left|\varphi_Y\left(\frac{x}{2}\right)\right|^2\left|\varphi_X\left(\frac{x}{2}\right)\right|^2 = |\varphi_X(x)|, \tag{2.1.2}$$

so

$$|\varphi_X(x)| = \left|\varphi_X\left(\frac{x}{2}\right)\right|^2\left|\varphi_Y\left(\frac{x}{2}\right)\right|^2 = \left|\varphi_X\left(\frac{x}{2}\right)\right|^2\left|\varphi_X\left(\frac{x}{2}\right)\right|^2 = \left|\varphi_X\left(\frac{x}{2}\right)\right|^4.$$

From this, we can conclude that φ_X is never zero, because if we assume there exists an $x \in \mathbb{R}$ such that $\varphi_X(x) = 0$, then

$$0 = |\varphi_X(x)| = \left|\varphi_X\left(\frac{x}{2}\right)\right|^4 = \left|\varphi_X\left(\frac{x}{4}\right)\right|^{16} = \left|\varphi_X\left(\frac{x}{8}\right)\right|^{64} = \ldots = \left|\varphi_X\left(\frac{x}{2^n}\right)\right|^{4^n} = \ldots,$$

so for all $n \in \mathbb{N} : \varphi_X\left(\frac{x}{2^n}\right) = 0$. But then, since φ_X is continuous,

$$\varphi_X(0) = \varphi_X\left(\lim_{n\to\infty}\frac{x}{2^n}\right) = \lim_{n\to\infty}\varphi_X\left(\frac{x}{2^n}\right) = 0,$$

however, $\varphi_X(0) = 1$ according to Proposition 1.5, which gives a contradiction. In the same way we can prove that φ_Y is never zero, ending the first part of the proof.

Part 2

In this part we will be looking at the following functions,

$$\Phi_X := \frac{\varphi_X}{|\varphi_X|}, \qquad \Phi_Y := \frac{\varphi_Y}{|\varphi_Y|},$$

and derive their general form. Notice how, because $|\varphi_X|$ and $|\varphi_Y|$ are never zero, these functions are defined for all $x \in \mathbb{R}$, and for the absolute value

$$|\Phi_X| = \left|\frac{\varphi_X}{|\varphi_X|}\right| = \frac{|\varphi_X|}{|\varphi_X|} = 1, \qquad |\Phi_Y| = \left|\frac{\varphi_Y}{|\varphi_Y|}\right| = \frac{|\varphi_Y|}{|\varphi_Y|} = 1. \tag{2.1.3}$$

Therefore Φ_X and Φ_Y are functions from $\mathbb{R}$ to $\mathbb{T}$. From (2.1.1) can be concluded that

$$\Phi_X(x)\Phi_Y(y) = \Phi_X\left(\frac{x+y}{2}\right)\Phi_Y\left(\frac{x+y}{2}\right)\Phi_X\left(\frac{x-y}{2}\right)\overline{\Phi_Y\left(\frac{x-y}{2}\right)}, \tag{2.1.4}$$

and when y is being substituted by $-y$ (with the 1^{st} statement of proposition 1.5 in mind),

$$\Phi_X(x)\overline{\Phi_Y(y)} = \Phi_X\left(\frac{x-y}{2}\right)\Phi_Y\left(\frac{x-y}{2}\right)\Phi_X\left(\frac{x+y}{2}\right)\overline{\Phi_Y\left(\frac{x+y}{2}\right)}. \tag{2.1.5}$$

Multiplying (2.1.4) and (2.1.5) will give

$$\Phi_X(x)^2\left|\Phi_Y(y)\right|^2 = \Phi_X\left(\frac{x+y}{2}\right)^2\left|\Phi_Y\left(\frac{x+y}{2}\right)\right|^2\Phi_X\left(\frac{x-y}{2}\right)^2\left|\Phi_Y\left(\frac{x-y}{2}\right)\right|^2,$$

and with (2.1.3) in mind we get

$$\Phi_X(x)^2 = \Phi_X\left(\frac{x+y}{2}\right)^2\Phi_X\left(\frac{x-y}{2}\right)^2.$$

Define the variables $u := \frac{x+y}{2}$ and $v := \frac{x-y}{2}$, then $u+v = x$ and therefore

$$\Phi_X(u+v)^2 = \Phi_X(u)^2\Phi_X(v)^2.$$

So Φ_X^2 is a homomorphism defined from $\mathbb{R}$ to $\mathbb{T}$. According to Lemma 2.2, this tells us that

$$\Phi_X(x)^2 = e^{i2\alpha x}$$

for some $\alpha \in \mathbb{R}$. This can be written as

$$\left(\Phi_X(x) + e^{i\alpha x}\right)\left(\Phi_X(x) - e^{i\alpha x}\right) = 0,$$

which is true for all $x \in \mathbb{R}$. Therefore, when we define the following sets

$$F_1 := \left\{x \in \mathbb{R} : \Phi_X(x) - e^{i\alpha x} = 0\right\}, \qquad F_2 := \left\{x \in \mathbb{R} : \Phi_X(x) + e^{i\alpha x} = 0\right\},$$

we know that $F_1 \cup F_2 = \mathbb{R}$. Since $e^{i\alpha x} \neq -e^{i\alpha x}$ for all $x \in \mathbb{R}$, F_1 and F_2 are disjoint. Furthermore, φ_X is a uniformly continuous function that is never zero, so we can assume that Φ_X is at least continuous. This means that F_1 and F_2 are closed sets, however, since they are disjoint and together they form $\mathbb{R}$, they also both have to be open. This is only possible if either $F_1 = \mathbb{R}$ or $F_2 = \mathbb{R}$. We already know $\Phi_X(0) = 1$ so $0 \in F_1$, which means $F_1 = \mathbb{R}$ and

$$\Phi_X(x) = e^{i\alpha x}. \tag{2.1.6}$$

Analogously,

$$\Phi_Y(x) = e^{i\beta x} \tag{2.1.6$'$}$$

for some $\beta \in \mathbb{R}$.

Part 3

Now we are going to calculate $|\varphi_X|$ and $|\varphi_Y|$. To do this, we will define the function

$$f(x) := \log|\varphi_X(x)|\,, \tag{2.1.7}$$

for all $x \in \mathbb{R}$. First of all, from (2.1.2) we know that

$$f(y) = \log|\varphi_Y(y)|\,, \tag{2.1.8}$$

and

$$f(-x) = \log|\varphi_X(-x)| = \log\left|\overline{\varphi_X(x)}\right| = \log|\varphi_X(x)| = f(x), \tag{2.1.9}$$

so $f(x)$ is symmetric. From (2.1.1) can be derived that

$$\log|\varphi_X(x)\varphi_Y(y)| = \log\left|\varphi_X\left(\frac{x+y}{2}\right)\varphi_Y\left(\frac{x+y}{2}\right)\varphi_X\left(\frac{x-y}{2}\right)\overline{\varphi_Y\left(\frac{x-y}{2}\right)}\right|.$$

Using the product rule of the logarithm and (2.1.8) on this equation leads to

$$f(x) + f(y) = 2\left[f\left(\frac{x+y}{2}\right) + f\left(\frac{x-y}{2}\right)\right],$$

or

$$f(2x) + f(2y) = 2\left[f\left(x+y\right) + f\left(x-y\right)\right]. \tag{2.1.10}$$

When we fill in $y = 0$ we get

$$f(2x) = 4f(x). \tag{2.1.11}$$

Now we will prove by induction that $f(kx) = k^2 f(x)$ for all $k \in \mathbb{N} \cup \{0\}$ and $x \in \mathbb{R}$. First we take $k = 0$,

$$f(0 \cdot x) = f(0) = \log|1| = 0 = 0^2 f(x),$$

so that holds. For $k = 1$ it is trivial and (2.1.11) shows that it is true for $k = 2$. Now we assume that it is true for $k = 0, 1, \ldots, n$ with $n \geq 2$. When we fill in $y = nx$ into (2.1.10) we see that

$$f(2x) + f(2nx) = 2\left[f((n+1)x) + f((n-1)x)\right].$$

Using our assumption and (2.1.11) gives us

$$4f(x) + 4f(nx) = 2f((n+1)x) + 2(n-1)^2 f(x).$$

We take the $f((n+1)x)$ apart and get

$$\begin{aligned} f((n+1)x) &= 2f(x) - (n-1)^2 f(x) + 2n^2 f(x) \\ &= \left(2 - (n-1)^2 + 2n^2\right) f(x), \end{aligned}$$

where

$$2 - (n-1)^2 + 2n^2 = 2 - (n^2 - 2n + 1) + 2n^2 = 1 + 2n + n^2 = (n+1)^2,$$

so

$$f((n+1)x) = (n+1)^2 f(x),$$

proving the property for all $k \in \mathbb{N} \cup \{0\}$. Since $f(x)$ is symmetric,

$$f(-kx) = f(kx) = k^2 f(x) = (-k)^2 f(x),$$

therefore this property is also true for all $k \in \mathbb{Z}$. And we can take it even further, take $k \in \mathbb{Z} \backslash \{0\}$, then

$$f(x) = f\left(k\frac{x}{k}\right) = k^2 f\left(\frac{x}{k}\right),$$

so

$$f\left(\frac{x}{k}\right) = \frac{1}{k^2} f(x).$$

Subsequently, when we take $r \in \mathbb{Q}$, we can write this as $r = \frac{m}{k}$ with $m \in \mathbb{Z}$ and $k \in \mathbb{N}$,

$$f(r) = f\left(\frac{m}{k}\right) = m^2 f\left(\frac{1}{k}\right) = \frac{m^2}{k^2} f(1) = r^2 f(1).$$

Since φ_X is uniformly continuous and never zero, we can again assume that f is at least continuous, therefore we can extend this property to the whole of $\mathbb{R}$, i.e. for all $x \in \mathbb{R}$,

$$f(x) = x^2 f(1).$$

Going back to the definition of f gives us

$$\log |\varphi_X(x)| = x^2 f(1),$$

which can be derived to

$$|\varphi_X(x)| = e^{-\frac{1}{2}\sigma^2 x^2}, \tag{2.1.12}$$

with $\frac{1}{2}\sigma^2 = -f(1)$ and due to Proposition 1.5 we see that

$$\frac{1}{2}\sigma^2 = -f(1) = -\log|\varphi_X(1)| \geq -\log|1| = 0.$$

From (2.1.8) we can also conclude that

$$|\varphi_Y(y)| = e^{-\frac{1}{2}\sigma^2 y^2}. \tag{2.1.12$'$}$$

Conclusion

From (2.1.6) we know that $\varphi_X(x) = e^{i\alpha x}|\varphi_X(x)|$ and when we fill in (2.1.12) we get

$$\varphi_X(x) = e^{i\alpha x - \frac{1}{2}\sigma^2 x^2}.$$

Doing the same with (2.1.6′) and (2.1.12′) gives us

$$\varphi_Y(y) = e^{i\beta x - \frac{1}{2}\sigma^2 x^2}$$

So according to our Definition 1.3, $X \sim N(\alpha, \sigma^2)$ and $Y \sim N(\beta, \sigma^2)$ therefore they are normally distributed, both with the same variance σ^2. □

2.2 Cramér's Theorem

In this section we will take a look at Cramér's theorem. The manner of proving this theorem comes from [5].

Theorem 2.3. **(Cramér's theorem)**
If X_1 and X_2 are independent random variables such that $X_1 + X_2$ has a normal distribution, then both X_1 and X_2 are normal.

To be able to prove this theorem, we will need to take a look at analytic characteristic functions. An analytic function is, simply put, a function that is differentiable in the complex plane. Since $\varphi_X(t)$ is only defined for real t, it needs to be extended to (a domain in) the complex plane. We shall call $\varphi_X(t)$ analytic on domain $D \subseteq C$ when this extension is analytic on D. Theorem 2.4 will look at a sufficient property for a characteristic function to be analytic.

Theorem 2.4.
If a random variable X has finite exponential moment $\mathbb{E}\left(e^{a|X|}\right) < \infty$ where $a > 0$, then its characterstic function $\varphi_X(t)$ is analytic in the strip $-a < \Im t < a$.

Proof. For the extension of $\varphi_X(t)$ to the complex plane, we go back to the definition of a characteristic function and extend it naturally as $\varphi_X(t) = \mathbb{E}\left(e^{itX}\right)$ with $t \in \mathbb{C}$. Writing $t = \alpha + i\beta$ with $\alpha, \beta \in \mathbb{R}$, we see that

$$\varphi_X(t) = \mathbb{E}\left(e^{i(\alpha+i\beta)X}\right) = \mathbb{E}\left(e^{i\alpha X}e^{-\beta X}\right),$$

therefore

$$\begin{aligned} |\varphi_X(t)| &= \left|\mathbb{E}\left(e^{i\alpha X}e^{-\beta X}\right)\right| \\ &\leq \mathbb{E}\left(\left|e^{i\alpha X}e^{-\beta X}\right|\right) \\ &= \mathbb{E}\left(\left|e^{-\beta X}\right|\right). \end{aligned}$$

Since e^x is non-negative and increasing, $|e^x| \leq e^{|x|}$ for all $x \in \mathbb{R}$, so

$$|\varphi_X(t)| \leq \mathbb{E}\left(e^{|\beta X|}\right),$$

which is finite as long as $|\beta| \leq a$ according to our assumption, so when $-a \leq \Im t \leq a$ we can be certain that our extension is well-defined.

Using the series expansion of the exponential function $e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}$ for our $\varphi_X(t)$, we will find that

$$\varphi_X(t) = \mathbb{E}\left(\sum_{n=0}^{\infty} \frac{i^n t^n X^n}{n!}\right) = \int_\Omega \sum_{n=0}^{\infty} \frac{i^n t^n X^n}{n!} dP.$$

On the strip $-a \leq \Im t \leq a$ we know that this is absolute convergent, so by Fubini's theorem we can interchange the integral and infinite sum to end up with

$$\sum_{n=0}^{\infty} \frac{i^n t^n \int_\Omega X^n dP}{n!} = \sum_{n=0}^{\infty} \frac{i^n t^n \mathbb{E}(X^n)}{n!}.$$

Notice how this series equals the Taylor series of φ_X around zero; $\sum_{n=0}^{\infty} \frac{t^n \varphi_X^{(n)}(0)}{n!}$. If a function equals its Taylors series in an open interval, than that function is analytic, so since φ_X is defined for $-a < \Im t < a$, it is also analytic in this strip. $\square$

This theorem has a pretty straightforward corollary, namely

Corollary 2.5.
If X is such that $\mathbb{E}\left(e^{a|X|}\right)$ for all $a > 0$, then its characteristic function $\varphi_X(t)$ is analytic in $\mathbb{C}$.

Before we go to the next lemma needed to prove Cramér's theorem, we will need a very specific upper bound for analytic functions. The Borel-Carathéodory theorem gives an upper bound that is dependent on the real part of the function and its value at zero, but not on the imaginary part, which will be very helpful.

Theorem 2.6. (Borel-Carathéodory theorem)
If a function f is analytic on a closed disc of radius R centered at the origin, then for $r < R$ the inequality

$$\sup_{|t|=r} |f(t)| \leq \frac{2r}{R-r} \sup_{|t|\leq R} \Re f(t) + \frac{R+r}{R-r}|f(0)|$$

holds.

Proof. If f is constant, this is trivial. Assume now that f is non-constant and $f(0) = 0$. Define $A(R) := \sup_{|t|\leq R} \Re f(t)$. Since $A(0) = 0$ we know that $A(R) > 0$, because if there exists an $R > 0$ such that $A(R) = 0$, then our function $\Re f$ attains a local maximum at 0 and must therefore be constant according to the maximum modulus principle, and because f is analytic, it must therefore also be constant. Furthermore, $\Re f(t) \leq A(R)$ for $|t| \leq R$ so on this closed disc, f sends every t to the half-plane P which is to the left of the line $x = A(R)$. In this proof we will make a function that maps P to a disc so we can apply Schwarz's Lemma [7, section III.3].

The function $\omega \mapsto \frac{\omega}{A(R)} - 1$ sends P to the left half plane, since

$$\Re\left(\frac{\omega}{A(R)} - 1\right) = \frac{\Re\omega}{A(R)} - 1 \leq \frac{A(R)}{A(R)} - 1 = 0,$$

and the function $\omega \mapsto R \cdot \frac{\omega+1}{\omega-1}$ sends the left half plane to the disc of radius R, since it is easy to see that if $\Re\omega \leq 0$ that $|\omega+1| \leq |\omega-1|$. Combining these two functions gives us

$$R \cdot \frac{\frac{\omega}{A(R)}}{\frac{\omega}{A(R)} - 2} = \frac{R\omega}{\omega - 2A(R)},$$

which apparently maps P to the disc of radius R. Since $f(t) < 2A(R)$ for every $|t| \leq R$, it is easy to see that

$$\frac{Rf(t)}{f(t) - 2A(R)}$$

is an analytic function in this closed disc. Schwarz's Lemma now states that for $|t| \leq R$

$$\left| \frac{Rf(t)}{f(t) - 2A(R)} \right| \leq |t|.$$

Take $|t| = r < R$, then

$$|Rf(t)| \leq r\,|f(t) - 2A(R)|\,,$$

so

$$|f(t)| \leq \frac{2r}{R-r} A(R), \tag{2.2.1}$$

which is the inequality from the theorem, given that $f(0) = 0$. If $f(0) \neq 0$, then we can use (2.2.1) on $f(t) - f(0)$

$$\begin{aligned} |f(t)| - |f(0)| &\leq |f(t) - f(0)| \\ &\leq \frac{2r}{R-r} \sup_{|t|\leq R} \Re\big(f(t) - f(0)\big) \\ &\leq \frac{2r}{R-r} \left(\sup_{|t|\leq R} \Re f(t) + |f(0)| \right). \end{aligned}$$

So when we bring $-|f(0)|$ to the other side we get

$$|f(t)| \leq \frac{2r}{R-r} A(r) + \left(\frac{2r}{R-r} + 1 \right) |f(0)| = \frac{2r}{R-r} A(R) + \frac{R+r}{R-r} |f(0)|.$$

□

Our final step before proving Cramér's theorem is the following lemma.

Lemma 2.7.
If X is a random variable such that $\mathbb{E}\left(e^{\lambda X^2}\right) < \infty$ for some $\lambda > 0$ and the complex extension of $\varphi_X(t)$ is analytic and never zero for all $t \in \mathbb{C}$, then X is normal.

Proof. Since our $\varphi_X(t)$ is never zero and it is continuous we can conclude that φ_X is positive, therefore we can define $g(t) := \log(\varphi_X(t))$ which is analytic for all $t \in \mathbb{C}$. Therefore we can write

$$g(t) = \sum_{n=0}^{\infty} a_n t^n,$$

where

$$a_n = \frac{1}{2\pi i} \oint_{C_r} \frac{g(z)}{z^{n+1}} dz,$$

with C_r the circle about 0 with radius $r > 0$. We want to prove that $g(t)$ is a quadratic polynomial by showing that $a_n = 0$ for $n > 2$. We will do this using the Borel-Carathéodory theorem. Since $g(0) = 0$ and $g(t)$ is analytic for all $t \in \mathbb{C}$, when we put $R = 2r$ we get that

$$\sup_{|t|=r} |g(t)| \leq 2 \sup_{|t|\leq 2r} \Re g(t). \tag{2.2.2}$$

So now we need an upper bound for $\Re g(t)$. When we write $t = a + bi$ and look at this real part we can see that

$$\Re g(t) = \log |\varphi_X(t)| = \log \left| \mathbb{E}\left(e^{i(a+bi)X}\right)\right| \leq \log \mathbb{E}\left|e^{i(a+bi)X}\right| \leq \log \mathbb{E} e^{|bX|}.$$

We can prove that in general $\mathbb{E}\left(e^{|bX|}\right) \leq Ce^{\frac{b^2}{2\lambda}}$ for a constant $C \in \mathbb{R}$, by verifying that the inequality $\lambda X^2 + \frac{b^2}{\lambda} \geq 2|bX|$ holds.

$$\lambda X^2 + \frac{b^2}{\lambda} - 2|bX| = \frac{\lambda^2 X^2 + b^2 - 2\lambda|bX|}{\lambda} = \frac{(\lambda|X| - |b|)^2}{\lambda},$$

so the inequality can be written as

$$\frac{(\lambda|X| - |b|)^2}{\lambda} \geq 0,$$

and since our $\lambda > 0$ this inequality holds for all X and b. Therefore

$$\mathbb{E}e^{|bX|} \leq \mathbb{E}\left(e^{\frac{\lambda X^2 + \frac{b^2}{\lambda}}{2}}\right) = Ce^{\frac{b^2}{2\lambda}},$$

with $C = \mathbb{E}\left(e^{\frac{\lambda X^2}{2}}\right)$, which is finite according to our assumption. In conclusion

$$\Re g(t) \leq \log C + \frac{b^2}{2\lambda}.$$

So according to (2.2.2) we find that

$$\sup_{|t|=r} |g(t)| \leq 2\log(C) + \frac{4r^2}{\lambda} = A + Br^2.$$

With $A, B \in \mathbb{R}$. So if $n > 2$ we have

$$|a_n| \leq \frac{1}{2\pi}\oint_{C_r} \frac{|g(z)|}{|z|^{n+1}} dz \leq \frac{1}{2\pi}\oint_{C_r} \frac{A + Br^2}{r^{n+1}} dz = \frac{1}{2\pi} \cdot \frac{A + Br^2}{r^{n+1}} \oint_{C_r} dz = \frac{A + Br^2}{r^n},$$

which converges to zero as $r \to \infty$, so we find that

$$g(t) = a_2 t^2 + a_1 t + a_0$$

and since $g(t) = \log(\varphi_X(t))$ it follows that

$$\varphi_X(t) = e^{a_2 t^2 + a_1 t + a_0}$$

which, according to Proposition 1.8, corresponds to a normal distribution. □

Proof of Cramér's theorem. Without loss of generality we can assume that the expectations of X_1 and X_2 are zero. To be able to use Lemma 2.7, we would need that $\mathbb{E}\left(e^{\lambda X_j^2}\right) < \infty$ for $j = 1, 2$ and that the characteristic functions are analytic and never zero. For the exponential moment, since $X_1 + X_2$ has a normal distribution, we know that

$$\mathbb{E}\left(e^{\lambda(X_1+X_2)^2}\right) < \infty,$$

for all $\lambda > 0$. Without loss of genereality, assume X_1 and X_2 are random variables on the probability triples $(\Omega_1, \mathcal{F}_1, P_1)$ and $(\Omega_2, \mathcal{F}_2, P_2)$ respectively. Now we define the functions $\hat{X}_1, \hat{X}_2 : \Omega_1 \times \Omega_2 \to \mathbb{R}$ as follows

$$\begin{aligned}\hat{X}_1(\omega_1, \omega_2) &= X(\omega_1),\\ \hat{X}_2(\omega_1, \omega_2) &= X(\omega_2).\end{aligned}$$

Define

$$\mathbb{E}_i(X) = \int_{\Omega_i} X dP_i$$

for $i = 1, 2$, then it follows that

$$\mathbb{E}_i(X_j) = \begin{cases} 0 & \text{if } i = j \\ X_j & \text{if } i \neq j \end{cases}$$

with $i, j \in \{1, 2\}$. So when we define $g(X) := e^{\lambda X^2}$,

$$\begin{aligned}\mathbb{E}\left(e^{\lambda X_1^2}\right) &= \mathbb{E}_1\left[g\left(X_1\right)\right]\\ &= \mathbb{E}_1\left[g\left(X_1 + \mathbb{E}_2(X_2)\right)\right]\\ &= \mathbb{E}_1\left[g\left(\mathbb{E}_2(X_1 + X_2)\right)\right],\end{aligned}$$

and since $g(X)$ is a convex function for some $\lambda > 0$, by Jensen's inequality we get

$$\mathbb{E}\left(e^{\lambda X_1^2}\right) \leq \mathbb{E}_1\mathbb{E}_2\left(g(X_1 + X_2)\right) = \mathbb{E}\left(e^{\lambda(X_1+X_2)^2}\right) < \infty.$$

In the same way can be shown that $\mathbb{E}\left(e^{\lambda X_2^2}\right) < \infty$. Since $x^2 \geq |x|$ for $|x| \geq 1$, we can conclude that $\mathbb{E}\left(e^{\lambda|X_j|}\right) < \infty$ for all $\lambda > 0$ and therefore, by Corollary 2.5, we know that the characteristic functions $\varphi_{X_1}(t)$ and $\varphi_{X_2}(t)$ are analytic in $\mathbb{C}$. Since

$$\varphi_{X_1}(t)\varphi_{X_2}(t) = \varphi_{X_1+X_2}(t) = e^{-\frac{\sigma^2 t}{2} + i\mu t}.$$

for some $\sigma, \mu \in \mathbb{R}$, neither characteristic functions can ever be zero. Therefore, Lemma 2.7 tells us that X_1 and X_2 are normal.

$\square$

2.3 Second characterization

In 1995, G. Bobkov and C. Houdré asked the question if a random variable X has a normal distribution if $\mathbb{P}\left(\left|\frac{X+Y}{\sqrt{2}}\right| > t\right) \leq \mathbb{P}(|X| > t)$ for $t > 0$, where X and Y are i.i.d. [4]. This statement has been proven by S. Kwapien, M. Pycia and W. Schachermayer in [8], but we can look at a more general characterization.

Theorem 2.8. (Second characterization)
Let $n \geq 2$ be a natural number. Let $a_1, a_2, ..., a_n$, with $a_i > 0$ for all i, be such that $\sum_{i=1}^n a_i^2 \geq 1$. If X, X_1, $X_2, \ldots,$ X_n are i.i.d. real random variables such that $\sum_{i=1}^n a_i X_i \overset{d}{=} X$ then X is normal.

Before we prove this characterization, we take a look at the following lemma where we also assume finite second moments.

Lemma 2.9.
Let $a_1, a_2, ..., a_n$, with $a_i > 0$ for all i, be such that and $\sum_{i=1}^n a_i^2 = 1$. If $X, X_1, X_2, \ldots, X_n$ are i.i.d. real random variables with finite second moment and $X \overset{d}{=} \sum_{i=1}^n a_i X_i$, then X is normal.

Proof. Since $X \overset{d}{=} \sum_{i=1}^n a_i X_i$ we know that

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i \mathbb{E}(X),$$

and because $\sum_{i=1}^n a_i > 1$, it follows that $\mathbb{E}(X) = 0$. For the characteristic function, according to (1.2.2) and the uniqueness theorem, the equality

$$\varphi_X(t) = \prod_{i=1}^n \varphi_X(a_i t) \tag{2.3.1}$$

holds for all $t \in \mathbb{R}$. This φ_X is never zero, because if there exists an $t \in \mathbb{R}$ where it is zero, then through (2.3.1) there has to be at least one a_i such that $\varphi_X(a_i t) = 0$. We can apply this multiple times and then we find that for general $s \in \mathbb{N}$ we have

$$\varphi_X(a_i^s t) = 0.$$

Since $a_i \in (0, 1)$ and our φ_X is continuous we find

$$0 = \lim_{s \to \infty} \varphi_X(a_i^s t) = \varphi_X(0),$$

which is in contradiction with Proposition 1.5. So our φ_X is never zero, and therefore we can define $g(t) := \log(\varphi_X(t))$ for all $t \in \mathbb{R}$. Our goal is to prove that $g(t)$ is a quadratic function $at^2 + bt + c$ because Proposition 1.8 then tells us that X is normally distributed. Because $X \overset{d}{=} \sum_{i=1}^n a_i X_i$ we know that

$$g(t) = \sum_{i_1=1}^n g(a_{i_1} t).$$

Applying the same property on $g(a_{i_1}t)$ we see that

$$g(t) = \sum_{i_1=1}^{n}\sum_{i_2=1}^{n} g(a_{i_1}a_{i_2}t),$$

and so for a general $m \in \mathbb{N}$ we find

$$g(t) = \sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_m=1}^{n} g\left(\prod_{j=1}^{m} a_{i_j}t\right).$$

The product within g can be written as $a_1^{k_1}a_2^{k_2}\ldots a_n^{k_n}$ where all the k_i are natural numbers with $k_1+k_2+\ldots+k_n = m$. All the summations in front of g make sure that we see every possibility for this product and a lot are being repeated multiple times. To understand how many times they are repeated we are going to need some combinatorics. Suppose all the k_i are given, then

$$\prod_{j=1}^{m} a_{i_j} = a_{i_1}a_{i_2}\ldots a_{i_m}$$

has k_1 times an a_1 in it, and we know that there are $\begin{pmatrix} m \\ k_1 \end{pmatrix} = \frac{m!}{k_1!(m-k_1)!}$ orders of the a_{i_j} where this happens. For a given order, there are still $m-k_1$ spots left, and there is k_2 times an a_2 in those empty spots, therefore there are $\begin{pmatrix} m-k_1 \\ k_2 \end{pmatrix} = \frac{(m-k_1)!}{k_2!(m-k_1-k_2)!}$ different ways to do this. We can repeat this for all the k_i and multiplying all these binomial coëfficiënts gives us

$$\frac{m!}{k_1!(m-k_1)!}\cdot\frac{(m-k_1)!}{k_2!(m-k_1-k_2)!}\cdot\ldots\cdot\frac{(k_{n-1}+k_n)!}{k_{n-1}!k_n!}\cdot 1 = \frac{m!}{k_1!k_2!\ldots k_{n-1}!k_n!}.$$

This is called the multinomial and is usually written as $\begin{pmatrix} m \\ k_1, k_2, \ldots, k_n \end{pmatrix}$. With this in our mind we can say that

$$g(t) = \sum_{k_1+k_2+\ldots+k_n=m}\begin{pmatrix} m \\ k_1, k_2, \ldots, k_n \end{pmatrix} g\left(\prod_{i=1}^{n} a_i^{k_i}t\right). \tag{2.3.2}$$

Now we will look at the second derivative of both sides

$$g''(t) = \sum_{k_1+k_2+\ldots+k_n=m}\begin{pmatrix} m \\ k_1, k_2, \ldots, k_n \end{pmatrix}\prod_{i=1}^{n} a_i^{2k_i} g''\left(\prod_{i=1}^{n} a_i^{k_i}t\right).$$

We want to show that this equals $g''(0)$ for all $t \in \mathbb{R}$ and is therefore constant. To do so, we will look at the difference of the two. But firstly, the multinomial theorem [3, p. 33] gives us that

$$\sum_{k_1+k_2+\ldots+k_n=m}\begin{pmatrix} m \\ k_1, k_2, \ldots, k_n \end{pmatrix}\prod_{i=1}^{n} a_i^{2k_i} = \left(\sum_{i=1}^{n} a_i^2\right)^m = 1,$$

therefore, when we choose an arbitrary $t_0 \in \mathbb{R}$,

$$\begin{aligned}\left|g''(t_0) - g''(0)\right| &= \left| \sum_{k_1+k_2+\ldots+k_n=m} \binom{m}{k_1, k_2, \ldots, k_n} \prod_{i=1}^{n} a_i^{2k_i} \left(g''\left(\prod_{i=1}^{n} a_i^{k_i} t_0\right) - g''(0)\right)\right| \\ &\leq \sum_{k_1+k_2+\ldots+k_n=m} \binom{m}{k_1, k_2, \ldots, k_n} \prod_{i=1}^{n} a_i^{2k_i} \left| g''\left(\prod_{i=1}^{n} a_i^{k_i} t_0\right) - g''(0)\right|.\end{aligned}$$

For every $m \in \mathbb{N}$ there is a combination of $k_{i,m}^*$ with $k_{1,m}^* + k_{2,m}^* + \ldots + k_{n,m}^* = m$ such that $\left|g''\left(\prod_{i=1}^{n} a_i^{k_{i,m}^*} t_0\right) - g''(0)\right| \geq \left|g''\left(\prod_{i=1}^{n} a_i^{k_i} t_0\right) - g''(0)\right|$ for all other combinations of k_i. That way

$$\begin{aligned}\left|g''(t_0) - g''(0)\right| &\leq \sum_{k_1+k_2+\ldots+k_n=m} \binom{m}{k_1, k_2, \ldots, k_n} \prod_{i=1}^{n} a_i^{2k_i} \left| g''\left(\prod_{i=1}^{n} a_i^{k_{i,m}^*} t_0\right) - g''(0)\right| \\ &= \left| g''\left(\prod_{i=1}^{n} a_i^{k_{i,m}^*} t_0\right) - g''(0)\right|.\end{aligned}$$

We know $\prod_{i=1}^{n} a_i^{k_{i,m}^*} t \in (0, a_1^m t_0)$ with $a_1^m < 1$. Now because $g(t) = \log(\varphi_X(t))$ we have

$$\begin{aligned}g'(t) &= \frac{\varphi_X'(t)}{\varphi_X(t)}, \\ g''(t) &= \frac{\varphi_X''(t)\varphi_X(t) - (\varphi_X'(t))^2}{\varphi_X(t)^2}.\end{aligned}$$

Since X has a finite second moment, and $\varphi_X(t)$ is never zero, it follows that $g''(t)$ is continuous, and even uniformly continuous on the compact interval $[0, t_0]$. Therefore, because $a_1^m t_0 \to 0$ when $m \to \infty$, we have that for all $\epsilon > 0$, there exists an $m \in \mathbb{N}$ such that for all $y \in (0, a_1^m t_0)$

$$\left|g''(y) - g''(0)\right| < \epsilon,$$

so we can conclude that

$$\left|g''(t_0) - g''(0)\right| \leq \lim_{m\to\infty} \left| g''\left(\prod_{i=1}^{n} a_i^{k_{i,m}^*} t_0\right) - g''(0)\right| = 0$$

for all $t_0 \in \mathbb{R}$, therefore $g''(t)$ is a constant, so

$$g(t) = at^2 + bt + c$$

for some $a, b, c \in \mathbb{C}$. Since $g(t) = \log|\varphi_X(t)|$ we get that

$$\varphi_X(t) = e^{at^2+bt+c}$$

which corresponds to a normal distribution according to Proposition 1.8. □

Although Lemma 2.9 looks very similar to our second characterization, we still assumed that all the random variables had finite second moments. This is a very strong condition and not needed. K. Oleskiewicz has written in his article [**4**] a more general, sufficient condition for X to be normally distributed as shown in the next theorem.

Theorem 2.10.
Let $n \geq 2$ be a natural number. Let $a_1 \geq a_2 \geq \cdots \geq a_n > 0$ be such that $\sum_{i=1}^n a_i^2 \geq 1$. If $X,\ X_1,\ X_2, \ldots,\ X_n$ are symmetric i.i.d. real random variables such that

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} a_i X_i\right| > t\right) \leq \mathbb{P}\left(|X| > t\right) \tag{2.3.3}$$

for any $t > 0$, then X has a normal distribution

Proof. First of all, when we put $\frac{a_i}{\sqrt{\sum_{i=1}^n a_i^2}}$ in place of a_i then (2.3.3) still holds. Therefore we can reduce this problem to the case of $\sum_{i=1}^n a_i^2 = 1$.

We define a function $h : \mathbb{R} \to [-1, 1]$ as follows:

$$h(x) = \begin{cases} \cos(x) & |x| < \pi, \\ -1 & |x| \geq \pi. \end{cases}$$

We are going to proof that the inequality

$$\frac{h(x+y) + h(x-y)}{2} \leq h(x)h(y) \tag{2.3.4}$$

holds for all $x, y \in \mathbb{R}$ by looking at three different cases.

Case 1: $|x|, |y| \geq \pi$
Since $h(x) \leq 1$ for all x, we know that

$$\frac{h(x+y) + h(x-y)}{2} \leq \frac{1+1}{2} = 1 = -1 \cdot -1 = h(x)h(y).$$

Case 2: $|x|, |y| < \pi$
In general $h(x) \leq \cos(x)$ for all x, therefore

$$\frac{h(x+y) + h(x-y)}{2} \leq \frac{\cos(x+y) + \cos(x-y)}{2}.$$

Using the trigonometrical equality $\cos(x \pm y) = \cos(x)\cos(y) \mp \sin(x)\sin(y)$ we see that

$$\frac{\cos(x+y) + \cos(x-y)}{2} = \cos(x)\cos(y) = h(x)h(y).$$

Case 3: $|x| < \pi$ and $|y| \geq \pi$
Let's assume that $y > 0$. In this case that means that $x+y \in (0, \infty)$ and $x-y \in (-\infty, 0)$. Because $h(x)$ is non-increasing on $\mathbb{R}_+$ and non-decreasing on $\mathbb{R}_-$ we can conclude that $h(x+y)$ and $h(x-y)$ are at their maximum when $y = \pi$ (if $y < 0$, the same would be true for $y = -\pi$). So

$$\frac{h(x+y) + h(x-y)}{2} \leq \frac{\cos(x+\pi) + \cos(x-\pi)}{2} = -\cos(x) = h(x)h(y).$$

Notice that because of the symmetry of $h(x)$ we know that $h(x-y) = h(y-x)$ and therefore the inequality is also true if $|y| < \pi$ and $|x| \geq \pi$.

We define another function $f_Y(s) := \mathbb{E}[h(sY)]$ for any random variable Y and $s \in \mathbb{R}$, however since $f_Y(0) = 1$ for any Y, there exists an s_0 such that for all i, $f_{X_i}(s) > 0$ for any $s \in (0, s_0]$. From now on we will only look at s in this interval. If Y and Z are both symmetric independent random variables, then (2.3.4) tells us that

$$f_Y(s)f_Z(s) = \mathbb{E}[h(sY)h(sZ)] \geq \mathbb{E}\left[\frac{h(s(Y+Z)) + h(s(Y-Z))}{2}\right] = \mathbb{E}[h(s(Y+Z))] = f_{Y+Z}(s),$$

and through basic induction,

$$f_{\sum\limits_{i=1}^{n} a_i X_i}(s) \leq \prod_{i=1}^{n} f_{a_i X_i}(s) = \prod_{i=1}^{n} f_X(a_i s). \tag{2.3.5}$$

Since $h(x)$ is symmetric we know $h(sX) = h(|sX|)$. Furthermore, since $h'(x) = -\sin(x)$ for $|x| < \pi$ and 0 everywhere else, $-h'(x) \geq 0$ for $x \geq 0$, therefore we can use Proposition 1.3 as follows,

$$f_X(s) = \mathbb{E}(h(sX)) = -\mathbb{E}(-h(|sX|)) = -\int_0^\infty \mathbb{P}\left(|X| > \frac{x}{|s|}\right)(-h'(x))dx + h(0).$$

Through (2.3.3) we find that

$$\begin{aligned} -\int_0^\infty \mathbb{P}\left(|X| > \frac{x}{|s|}\right)(-h'(x))dx + h(0) &\leq -\int_0^\infty \mathbb{P}\left(\left|\sum_{i=1}^{n} a_i X_i\right| > \frac{x}{|s|}\right)(-h'(x))dx + h(0) \\ &= f_{\sum\limits_{i=1}^{n} a_i X_i}(s), \end{aligned}$$

so with (2.3.5) in our mind we get

$$f_X(s) \leq f_{\sum\limits_{i=1}^{n} a_i X_i}(s) \leq \prod_{i=1}^{n} f_X(a_i s). \tag{2.3.6}$$

Let $g(s) = f_X(s)^{\frac{1}{s^2}}$, then through (2.3.6) we get

$$\prod_{i=1}^{n} g(a_i s)^{a_i^2} = \prod_{i=1}^{n} f_X(a_i s)^{\frac{1}{s^2}} = \left(\prod_{i=1}^{n} f_X(a_i s)\right)^{\frac{1}{s^2}} \geq f_X(s)^{\frac{1}{s^2}} = g(s),$$

and so for $s > 0$,

$$g(s) \leq \prod_{i=1}^{n} g(a_i s)^{a_i^2} \leq \sup_{r\in(0,a_1 s]} \prod_{i=1}^{n} g(r)^{a_1^2} = \sup_{r\in(0,a_1 s]} g(r)^{\sum_{i=1}^{n} a_i^2} = \sup_{r\in(0,a_1 s]} g(r). \tag{2.3.7}$$

Through induction we can then find that for general $m \in \mathbb{N}$ we have

$$g(s) \leq \sup_{r\in(0,a_1^m s]} g(r),$$

and since $a_1 < 1$ we know that $a_1^m s$ goes to zero if m goes to infinity. Therefore, since we can write $g(s_0) = e^{-2c} > 0$ with $c \in \mathbb{R}_+$, we see that

$$\limsup_{r\to 0} g(r) \geq g(s_0) = e^{-2c}.$$

Define a sequence as follows; we take $r_1 := s_0$ and then we choose our r_k with $k \geq 2$ such that $0 < r_k \leq a_1 r_{k-1}$ and $g(r_k) \geq g(a_1 r_{k-1})$, which exists according to (2.3.7). This way, $r_k \to 0$ and $g(r_k) \geq e^{-2c}$. Now

$$\mathbb{E}(\cos(r_k X)) \geq \mathbb{E}(h(r_k X)) = f_X(r_k) \geq e^{-2cr_k^2} \geq 1 - 2cr_k^2,$$

or

$$c \geq \frac{1}{r_k^2}\mathbb{E}\left[\frac{1-\cos(r_k X)}{2}\right].$$

In general, the equality $\sin^2\left(\frac{x}{2}\right) = \sqrt{\frac{1-\cos(2x)}{2}}$ holds, so in our case

$$c \geq \frac{1}{r_k^2}\mathbb{E}\left[\sin^2\left(\frac{r_k X}{2}\right)\right],$$

and since $\sin(x) \geq \frac{x}{\pi}$ for $0 \leq x \leq \pi$ we get that

$$c \geq \frac{1}{r_k^2}\mathbb{E}\left[\left(\frac{r_k X}{\pi}\right)^2\right]\mathbb{1}_{|r_k X|\leq\pi} = \mathbb{E}\left[\left(\frac{X}{\pi}\right)^2\right]\mathbb{1}_{|r_k X|\leq\pi}.$$

Because $r_k \to 0$, we know that $\mathbb{1}_{|r_k X|\leq\pi} \to 1$, so

$$c \geq \frac{1}{\pi^2}\mathbb{E}(X^2).$$

In conclusion, X has a finite second moment. Now we will take a look at $\mathbb{E}\left(\sum_{i=1}^n a_i X_i\right)^2$.

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} a_i X_i\right)^2\right] = \mathbb{E}\left[\sum_{i=1}^{n} a_i^2 X_i^2 + \sum_{i\neq j} a_i a_j X_i X_j\right] = \sum_{i=1}^{n} a_i^2 \mathbb{E}\left(X_i^2\right) + \sum_{i\neq j} a_i a_j \mathbb{E}(X_i X_j).$$

Since all the X_i are independent, $\mathbb{E}(X_i X_j) = \mathbb{E}(X_i)\mathbb{E}(X_j)$, and they all are i.i.d. with zero expectation, so

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} a_i X_i\right)^2\right] = \sum_{i=1}^{n} a_i^2 \mathbb{E}(X^2) = \mathbb{E}(X^2). \tag{2.3.8}$$

With Proposition 1.3 and (2.3.3), it follows that

$$\mathbb{E}\left[\left|\sum_{i=1}^{n} a_i X_i\right|^2\right] = \int_0^\infty \mathbb{P}\left(\left|\sum_{i=1}^{n} a_i X_i\right| > t\right) 2tdt \leq \int_0^\infty \mathbb{P}(|X| > t)2tdt = \mathbb{E}(|X|^2),$$

so by combining this with (2.3.8) we can see that

$$\int_0^\infty \mathbb{P}\left(\left|\sum_{i=1}^{n} a_i X_i\right| > t\right) 2tdt = \int_0^\infty \mathbb{P}(|X| > t)2tdt. \tag{2.3.9}$$

Now, if there was a $t_0 > 0$ such that $\mathbb{P}\left(\left|\sum_{i=1}^{n} a_i X_i\right| > t_0\right) < \mathbb{P}(|X| > t_0)$ then, because of the right-continuity of the probabilities, there would exist an $\epsilon > 0$ such that $\mathbb{P}\left(\left|\sum_{i=1}^{n} a_i X_i\right| > t\right) < \mathbb{P}(|X| > t)$ for $t \in [t_0, t_0 + \epsilon)$, and then

$$\int_0^\infty \mathbb{P}\left(\left|\sum_{i=1}^{n} a_i X_i\right| > t\right) 2tdt < \int_0^\infty \mathbb{P}(|X| > t)2tdt.$$

Which contradicts (2.3.9), so $\mathbb{P}\left(\left|\sum_{i=1}^{n} a_i X_i\right| > t\right) = \mathbb{P}(|X| > t)$ for $t > 0$. Because X and all the X_i are symmetric it follows that

$$\mathbb{P}\left(\sum_{i=1}^{n} a_i X_i > t\right) = \mathbb{P}(X > t).$$

For general Y we have $\mathbb{P}(Y > t) = 1 - \mathbb{P}(Y \leq t)$, therefore

$$\mathbb{P}\left(\sum_{i=1}^{n} a_i X_i \leq t\right) = \mathbb{P}(X \leq t),$$

and by using symmetry again, we see that

$$\mathbb{P}\left(\sum_{i=1}^{n} a_i X_i \geq -t\right) = \mathbb{P}(X \geq -t)$$

for all $t > 0$, and hence for all $t \in \mathbb{R}$. Therefore $\sum_{i=1}^{n} a_i X_i \stackrel{d}{=} X$, so by Lemma 2.9, X has a normal distribution. $\square$

With this theorem, the proof of the second characterization becomes fairly easy.

Proof of theorem 2.8. Let X' be an independent copy of X, then there also exist $X_1', X_2', ..., X_n'$ i.i.d. such that $\sum_{i=1}^n a_i X_i' \stackrel{d}{=} X'$, therefore

$$X - X' \stackrel{d}{=} \sum_{i=1}^n a_i X_i - \sum_{i=1}^n a_i X_i' = \sum_{i=1}^n a_i (X_i - X_i').$$

Since all the $X_i - X_i'$ are symmetric, Theorem 2.10 now tells us that $X - X'$ has a normal distribution. By applying Cramér's theorem we find that X has a normal distribution as well. □

Chapter 3

Gaussian measures on abstract spaces

In the previous chapter we have seen certain characterizations of a normal distribution on $\mathbb{R}$. In this section we will turn some of these characterizations into definitions in order to define them on different spaces than just $\mathbb{R}$. The information in this chapter is derived from the lecture notes of W. Bryc [5].

3.1 Definitions

A special case of our second characterization, which is often examined, is the case where $n = 2$ and $a_1 = a_2 = \frac{1}{\sqrt{2}}$. In this case we define the Equidistributed-Gaussian, or $\mathcal{E}$-Gaussian, random variables. Since this theorem only uses addition and scalar multiplication, we won't necessarily need the structure of $\mathbb{R}^d$, but any vector space V would do. But first we will need to give a definition of a random variable on such a vector space.

Suppose V is a vector space over the field $\mathbb{R}$ which has a σ-algebra $\mathcal{F}$ (if we are working with a topological space, the obvious choice for $\mathcal{F}$ would be the Borel σ-algebra) such that scalar multiplication $(\mathbf{v}, t) \mapsto t\mathbf{v}$ and vector addition $(\mathbf{v}, \mathbf{w}) \mapsto \mathbf{v} + \mathbf{w}$ are measurable transformations $V \times \mathbb{R} \to V$ and $V \times V \to V$ with respect to the σ-fields $\mathcal{F} \otimes \mathcal{B}_{\mathbb{R}}$ and $\mathcal{F} \otimes \mathcal{F}$ respectively. If $(\Omega, \mathcal{M}, \mathbb{P})$ is a probability space, then a measurable function $\mathbf{X} : \Omega \to V$ is called a V-valued random variable.

Definition 3.1.
Let V be a vector space. A V-valued random variable $\mathbf{X}$ is $\mathcal{E}$-Gaussian if the distribution of $\sqrt{2}\mathbf{X}$ is equal to the distribution of $\mathbf{X} + \mathbf{X}'$, where *$\mathbf{X}'$ is an independent copy of $\mathbf{X}$.*

One way to make an $\mathcal{E}$-Gaussian random variable is to take the product of an $\mathbb{R}$-valued normal random variable $X \sim N(m, \sigma^2)$ and a vector $\mathbf{v} \in V$. This has to do with the fact that $\sqrt{2}X \stackrel{d}{=} X + X'$, for X' an independent copy. Because of this,

$$\mathbb{P}(\sqrt{2}X\mathbf{v} \in B) = \mathbb{P}(\sqrt{2}X \in C) = \mathbb{P}(X + X' \in C) = \mathbb{P}((X + X')\mathbf{v} \in B)$$

for all $B \in \mathcal{F}$ and $C := \{x \in \mathbb{R} : x\mathbf{v} \in B\}$, therefore $\sqrt{2}X\mathbf{v} \stackrel{d}{=} (X + X')\mathbf{v}$.

The next definition is derived from Theorem 2.1. Since it determines whether a random variable is normal by using independence, we will call them Independent-Gaussian, or $\mathcal{I}$-Gaussian. This theorem only uses the basic operations $+$ and $-$, therefore we can look at very basic structures that only use these operations, like an abelian group G.

Let G be an abelian group with σ-algebra $\mathcal{F}$ such that the group operation $\mathbf{v}, \mathbf{w} \mapsto \mathbf{v}+\mathbf{w}$ is a measurable transformation $G \times G \to G$. $\mathbf{v} - \mathbf{w}$ is also measurable, as it is equal to $\mathbf{v} + \mathbf{w}^\dagger$, where $\mathbf{w}^\dagger$ is the additive inverse of $\mathbf{w}$ which is also an element of the group G. If $(\Omega, \mathcal{M}, \mathbb{P})$ is a probability space, then a measurable function $\mathbf{X} : \Omega \to G$ is a G-valued random variable.

Definition 3.2.

Let G be an abelian group. A G-valued random variable $\mathbf{X}$ is $\mathcal{I}$-Gaussian if the random variables $\mathbf{X}+\mathbf{X}'$ and $\mathbf{X}-\mathbf{X}'$, where $\mathbf{X}'$ is an independent copy of $\mathbf{X}$, are independent.

An example of an $\mathcal{I}$-Gaussian can be made with the help of a measurable homomorphism $\phi : \mathbb{R} \to G$. Take a normal random variable X on $\mathbb{R}$, then $\phi(X)$ is $\mathcal{I}$-Gaussian. Namely, since X is normally distributed we know that $X+X'$ and $X-X'$ are independent, where X' is an independent copy. Since ϕ is a measurable function, it preserves independence, so $\phi(X+X')$ and $\phi(X-X')$ are also independent. After using the property of homomorphisms, the same applies to $\phi(X)+\phi(X')$ and $\phi(X)-\phi(X')$, therefore $\phi(X)$ is $\mathcal{I}$-Gaussian.

A multivariate normal random variable, as defined in Definition 1.4, can also be defined on other spaces than just $\mathbb{R}^d$. To do this, we will need to work with linear forms.

Definition 3.3.

A function $f : V \to F$, with V a vector space and F a field, is called a linear form if

1. $f(\mathbf{v}+\mathbf{w}) = f(\mathbf{v}) + f(\mathbf{w})$ *for all* $\mathbf{v}, \mathbf{w} \in V$
2. $f(a\mathbf{v}) = af(\mathbf{v})$ *for all* $\mathbf{v} \in V, a \in F$

We will only be looking at the case where $F = \mathbb{R}$. Now we are able to define multivariate normal random variables on other spaces, which we will call Linear-Gaussian, or $\mathcal{L}$-Gaussian for short, since it uses linear forms to determine whether a random variable is normal or not.

Definition 3.4.

Let V be a vector space with topology τ. A V_τ-valued random vector $\mathbf{X}$ is $\mathcal{L}$-Gaussian if for every continuous linear form f the real valued random variable $f(\mathbf{X})$ is normal.

Note that if $V = \mathbb{R}^d$, then this definition is equivalent with Definition 1.4, where the linear forms are the inner products with a vector $\mathbf{t} \in \mathbb{R}^d$.

3.2 Equivalence of the Gaussians

We now have defined three different types of a Gaussian. It isn't surprising that a lot of times these three definitions are equivalent, however there are times when they aren't ([11] gives examples that are $\mathcal{E}$-Gaussian without being $\mathcal{I}$-Gaussian). One situation where we are certain that they are equivalent is when we look at centered random variables on $\mathbb{R}$, since we have derived our definitions from characterizations of a normal distribution in $\mathbb{R}$. But there is still room for expansion.

Proposition 3.1.
If $\mathbf{X}$ is a symmetric random variable on $\mathbb{R}^d$, then the following statements are equivalent:

- $\mathbf{X}$ *is $\mathcal{L}$-Gaussian.*
- $\mathbf{X}$ *is $\mathcal{E}$-Gaussian.*
- $\mathbf{X}$ *is $\mathcal{I}$-Gaussian.*

Before we prove this we will first need the following lemma.

Lemma 3.2.
Let $\mathbf{X}, \mathbf{Y} : \Omega \to \mathbb{R}^d$ be random vectors, and $f : \mathbb{R}^d \to \mathbb{R}$ a borel-measurable function. If $\mathbf{X} \overset{d}{=} \mathbf{Y}$, then $f(\mathbf{X}) \overset{d}{=} f(\mathbf{Y})$

Proof. The proof of this is rather straightforward. Since $\mathbf{X} \overset{d}{=} \mathbf{Y}$ we know that

$$\mathbb{P}(\mathbf{X} \in A) = \mathbb{P}(\mathbf{Y} \in A)$$

for all $A \in \mathcal{B}(\mathbb{R}^d)$. Now take $B \in \mathcal{B}(\mathbb{R})$, then

$$\mathbb{P}(f(\mathbf{X}) \in B) = \mathbb{P}(\mathbf{X} \in f^{-1}(B)),$$

where $f^{-1}(B) = \{x \in \mathbb{R}^d : f(x) \in B\}$. Since f is borel-measurable, we know that $f^{-1}(B) \in \mathcal{B}(\mathbb{R}^d)$, therefore

$$\mathbb{P}(\mathbf{X} \in f^{-1}(B)) = \mathbb{P}(\mathbf{Y} \in f^{-1}(B)) = \mathbb{P}(f(\mathbf{Y}) \in B),$$

so $f(\mathbf{X}) \overset{d}{=} f(\mathbf{Y})$. □

Proof of proposition 3.1. We will prove this by showing that if $\mathbf{X}$ is $\mathcal{E}$-Gaussian or $\mathcal{I}$-Gaussian, then it is also $\mathcal{L}$-Gaussian, and if it is $\mathcal{L}$-Gaussian then it is $\mathcal{E}$-Gaussian and $\mathcal{I}$-Gaussian.

<u>$\mathcal{E}$-Gaussian $\Rightarrow$ $\mathcal{L}$-Gaussian</u>
If $\sqrt{2}\mathbf{X} \overset{d}{=} \mathbf{X} + \mathbf{X}'$, then Lemma 3.2 tells us that for an arbitrary $\mathbf{t} \in \mathbb{R}^d$

$$\sqrt{2}\mathbf{t}^T\mathbf{X} \overset{d}{=} \mathbf{t}^T\mathbf{X} + \mathbf{t}^T\mathbf{X}'.$$

Using Theorem 2.8 on $\mathbf{t}^T\mathbf{X}$ tells us that it is normally distributed, and therefore $\mathbf{X}$ is $\mathcal{L}$-Gaussian. This proof also extends to general vector spaces V.

$\mathcal{I}$-Gaussian $\Rightarrow$ $\mathcal{L}$-Gaussian
If $\mathbf{X}+\mathbf{X}'$ and $\mathbf{X}-\mathbf{X}'$ are independent, then it follows that $\mathbf{t}^T\mathbf{X}+\mathbf{t}^T\mathbf{X}'$ and $\mathbf{t}^T\mathbf{X}-\mathbf{t}^T\mathbf{X}'$ are also independent. So Theorem 2.1 now states that $\mathbf{t}^T\mathbf{X}$ is normal, therefore $\mathbf{X}$ is $\mathcal{L}$-Gaussian. Again, this proof extends to general vector spaces V.

$\mathcal{L}$-Gaussian $\Rightarrow$ $\mathcal{E}$-Gaussian
Assume $\mathbf{X}$ is a centered $\mathcal{L}$-Gaussian witch covariance matrix Σ, and $\mathbf{X}'$ is an independent copy. From Proposition 1.9 we know that $\mathbf{X}+\mathbf{X}' \sim \mathcal{N}(0, 2\Sigma)$. To find the distribution of $\sqrt{2}\mathbf{X}$ we turn to the characteristic function.

$$\varphi_{\sqrt{2}\mathbf{X}}(\mathbf{t}) = \varphi_{\mathbf{X}}(\sqrt{2}\mathbf{t}) = \exp\left(-\frac{1}{2}\sqrt{2}\mathbf{t}^T\Sigma\sqrt{2}\mathbf{t}\right) = \exp\left(-\mathbf{t}^T\Sigma\mathbf{t}\right).$$

From the characteristic function we can see that $\sqrt{2}\mathbf{X} \sim \mathcal{N}(0, 2\Sigma)$, hence $\sqrt{2}\mathbf{X} \stackrel{d}{=} \mathbf{X}+\mathbf{X}'$.

$\mathcal{L}$-Gaussian $\Rightarrow$ $\mathcal{I}$-Gaussian
This proof will work similar as the proof of the first implication of theorem 2.1.1, but now using the characteristic function of an $\mathcal{L}$-Gaussian, given by (1.4.2). Assume $\mathbf{X}$ is a centered $\mathcal{L}$-Gaussian with covariance matrix Σ and let $\mathbf{X}'$ be an independent copy. Still due to the independence we can find that

$$\varphi_{(\mathbf{X}+\mathbf{X}',\mathbf{X}-\mathbf{X}')}(\mathbf{s},\mathbf{t}) = \varphi_{\mathbf{X}}(\mathbf{s}+\mathbf{t}) \cdot \varphi_{\mathbf{X}'}(\mathbf{s}-\mathbf{t}),$$

and

$$\varphi_{\mathbf{X}}(\mathbf{s}+\mathbf{t}) \cdot \varphi_{\mathbf{X}'}(\mathbf{s}-\mathbf{t}) = \exp\left(-\frac{1}{2}(\mathbf{s}+\mathbf{t})^T\Sigma(\mathbf{s}+\mathbf{t}) - \frac{1}{2}(\mathbf{s}-\mathbf{t})^T\Sigma(\mathbf{s}-\mathbf{t})\right).$$

After removing all the brackets and simplifying we are left with

$$\begin{aligned}\varphi_{\mathbf{X}}(\mathbf{s}+\mathbf{t}) \cdot \varphi_{\mathbf{X}'}(\mathbf{s}-\mathbf{t}) &= \exp\left(-\frac{1}{2}\left(\mathbf{s}^T\Sigma\mathbf{s} + \mathbf{t}^T\Sigma\mathbf{t} + \mathbf{s}^T\Sigma\mathbf{s} + \mathbf{t}^T\Sigma\mathbf{t}\right)\right) \\ &= \exp\left(-\mathbf{s}^T\Sigma\mathbf{s}\right) \cdot \exp\left(-\mathbf{t}^T\Sigma\mathbf{t}\right) \\ &= \varphi_{\mathbf{X}+\mathbf{X}'}(\mathbf{s}) \cdot \varphi_{\mathbf{X}-\mathbf{X}'}(\mathbf{t}).\end{aligned}$$

So according to Proposition 1.7 we find that $\mathbf{X}+\mathbf{X}'$ and $\mathbf{X}-\mathbf{X}'$ are independent, therefore $\mathbf{X}$ is $\mathcal{I}$-Gaussian. $\square$

So $\mathbb{R}^d$ still has enough structure for these three definitions to be equivalent. But there are other spaces with enough structure, for example $C[0,1]$. The following theorem gives us a sufficient condition for V to have this property.

Theorem 3.3.
If $\boldsymbol{X}$ is a symmetric random variable on V with V a separable Banach space, then the statements from Proposition 3.1 are equivalent.

Here seperable means that it contains a countable, dense subset, and a Banach space is a complete normed vector space. The proof of this theorem goes further into the material than desired, and will therefore not be included in this report.

3.3 Integrability of an $\mathcal{I}$-Gaussian

Much like a normal random variable, all the moments of an $\mathcal{I}$-Gaussian exist. We will do this by proving that it has exponential integrability properties, which we can derive from the following theorem.

Theorem 3.4.
Let $\mathbf{X}$ be $\mathcal{I}$-Gaussian on a measurable Abelian group G. Let $d : G \to \mathbb{R}^+$ be a measurable function such that $d(a+b) \leq d(a) + d(b)$ and $d(-a) = d(a)$ for every $a, b \in G$. Suppose that for some $\delta \in \left(0, \frac{1}{2}\right)$ and $\eta > 0$

$$\mathbb{P}(d(\mathbf{X}) \geq \eta) \leq \delta, \tag{3.3.1}$$

then

$$\mathbb{P}(d(\mathbf{X}) \geq t) \leq \exp\left(\frac{t^2}{144\eta^2} \log \frac{\delta}{1-\delta}\right) \tag{3.3.2}$$

for every $t \geq 3\eta$.

Proof. Let $\mathbf{Y}$ be an independent copy of $\mathbf{X}$. We shall first prove that the inequality

$$\mathbb{P}(d(\mathbf{X}) < \eta)^3 \mathbb{P}(d(\mathbf{X}) \geq t) \leq \mathbb{P}\left(d(\mathbf{X}) \geq \frac{t - 3\eta}{2}\right)^4 \tag{3.3.3}$$

holds. Since $\mathbf{Y}$ and $\mathbf{X}$ are i.i.d., we see that

$$\begin{aligned}\mathbb{P}(d(\mathbf{X}) < \eta)^3 \mathbb{P}(d(\mathbf{X}) \geq t) &= \mathbb{P}(d(\mathbf{X}) < \eta)\mathbb{P}(d(\mathbf{Y}) < \eta)\mathbb{P}(d(\mathbf{X}) \geq t)\mathbb{P}(d(\mathbf{Y}) < \eta) \\ &= \mathbb{P}(d(\mathbf{X}) < \eta, d(\mathbf{Y}) < \eta)\mathbb{P}(d(\mathbf{X}) \geq t, d(\mathbf{Y}) < \eta).\end{aligned}$$

Using the properties of d from our assumptions, it follows that $d(\mathbf{X}) < \eta$ and $d(\mathbf{Y}) < \eta$ always implies the following,

$$\begin{aligned}d(\mathbf{X} + \mathbf{Y}) &\leq d(\mathbf{X}) + d(\mathbf{Y}) < 2\eta, \\ d(\mathbf{X} - \mathbf{Y}) &\leq d(\mathbf{X}) + d(-\mathbf{Y}) < 2\eta.\end{aligned}$$

So since one event implies the other, the probability of that one event has to be smaller than the other one, therefore

$$\mathbb{P}(d(\mathbf{X}) < \eta, d(\mathbf{Y}) < \eta) \leq \mathbb{P}(d(\mathbf{X} + \mathbf{Y}) < 2\eta, d(\mathbf{X} - \mathbf{Y}) < 2\eta).$$

In the same way we can find that

$$\mathbb{P}(d(\mathbf{X}) \geq t, d(\mathbf{Y}) < \eta) \leq \mathbb{P}(d(\mathbf{X} + \mathbf{Y}) \geq t - \eta, d(\mathbf{X} - \mathbf{Y}) \geq t - \eta),$$

so

$$\begin{aligned}&\mathbb{P}(d(\mathbf{X}) < \eta)^3 \mathbb{P}(d(\mathbf{X}) \geq t) \\ &\quad \leq \mathbb{P}(d(\mathbf{X} + \mathbf{Y}) < 2\eta, d(\mathbf{X} - \mathbf{Y}) < 2\eta)\mathbb{P}(d(\mathbf{X} + \mathbf{Y}) \geq t - \eta, d(\mathbf{X} - \mathbf{Y}) \geq t - \eta) \\ &\quad = \mathbb{P}(d(\mathbf{X} + \mathbf{Y}) < 2\eta)\mathbb{P}(d(\mathbf{X} - \mathbf{Y}) < 2\eta)\mathbb{P}(d(\mathbf{X} + \mathbf{Y}) \geq t - \eta)\mathbb{P}(d(\mathbf{X} - \mathbf{Y}) \geq t - \eta) \\ &\quad = \mathbb{P}(d(\mathbf{X} + \mathbf{Y}) < 2\eta, d(\mathbf{X} - \mathbf{Y}) \geq t - \eta)\mathbb{P}(d(\mathbf{X} + \mathbf{Y}) \geq t - \eta, d(\mathbf{X} - \mathbf{Y}) < 2\eta),\end{aligned}$$

with the second and third row using the independence of $\mathbf{X}+\mathbf{Y}$ and $\mathbf{X}-\mathbf{Y}$. Using the fact that $d(a) \geq d(a+b) - d(b)$ for $a, b \in G$, if we would write $a = 2\mathbf{X}$ and $b = -\mathbf{X}+\mathbf{Y}$, we would find that $d(2\mathbf{X}) \geq d(\mathbf{X}+\mathbf{Y}) - d(\mathbf{X}-\mathbf{Y})$. So from $d(\mathbf{X}+\mathbf{Y}) < 2\eta$ and $d(\mathbf{X}-\mathbf{Y}) \geq t-\eta$ it follows that

$$d(2\mathbf{X}) \geq d(\mathbf{X}+\mathbf{Y}) - d(-\mathbf{X}+\mathbf{Y}) \geq t - 3\eta,$$

and similarly

$$d(2\mathbf{Y}) \geq d(\mathbf{X}+\mathbf{Y}) - d(\mathbf{X}-\mathbf{Y}) \geq t - 3\eta.$$

So using the same method as before, we find that

$$\mathbb{P}(d(\mathbf{X}+\mathbf{Y}) < 2\eta, d(\mathbf{X}-\mathbf{Y}) \geq t-\eta) \leq \mathbb{P}(d(2\mathbf{X}) \geq t-3\eta, d(2\mathbf{Y}) \geq t-3\eta),$$

and again, in the same way we can find that

$$\mathbb{P}(d(\mathbf{X}+\mathbf{Y}) \geq t-\eta, d(\mathbf{X}-\mathbf{Y}) < 2\eta) \leq \mathbb{P}(d(2\mathbf{X}) \geq t-3\eta, d(2\mathbf{Y}) \geq t-3\eta).$$

Therefore

$$\begin{aligned}
&\mathbb{P}(d(\mathbf{X}) < \eta)^3 \mathbb{P}(d(\mathbf{X}) \geq t) \\
&\qquad \leq \mathbb{P}(d(2\mathbf{X}) \geq t-3\eta, d(2\mathbf{Y}) \geq t-3\eta)\mathbb{P}(d(2\mathbf{X}) \geq t-3\eta, d(2\mathbf{Y}) \geq t-3\eta) \\
&\qquad = \mathbb{P}(d(2\mathbf{X}) \geq t-3\eta)^4 \\
&\qquad \leq \mathbb{P}\left(d(\mathbf{X}) \geq \frac{t-3\eta}{2}\right)^4.
\end{aligned}$$

So we got our inequality (3.3.3), and by (3.3.1) we also know that $\mathbb{P}(d(\mathbf{X}) < \eta) \geq 1-\delta$, therefore

$$(1-\delta)^3 \mathbb{P}(d(\mathbf{X}) \geq t) \leq \mathbb{P}\left(d(\mathbf{X}) \geq \frac{t-3\eta}{2}\right)^4.$$

When we define $t_n := (2^{n+1}-1)3\eta$ for $n \geq 0$, then

$$\frac{t_n - 3\eta}{2} = \frac{2^{n+1} \cdot 3\eta - 6\eta}{2} = (2^n - 1)3\eta = t_{n-1},$$

so we find that

$$\mathbb{P}(d(\mathbf{X}) \geq t_n) \leq (1-\delta)^{-3} \mathbb{P}\left(d(\mathbf{X}) \geq t_{n-1}\right)^4.$$

Using the same inequality again on $\mathbb{P}\left(d(\mathbf{X}) \geq t_{n-1}\right)$ gives us

$$\begin{aligned}
\mathbb{P}(d(\mathbf{X}) \geq t_n) &\leq (1-\delta)^{-3}\left((1-\delta)^{-3}\mathbb{P}\left(d(\mathbf{X}) \geq t_{n-2}\right)^4\right)^4 \\
&= (1-\delta)^{-3(1+4)}\mathbb{P}\left(d(\mathbf{X}) \geq t_{n-2}\right)^{4^2}
\end{aligned}$$

and again on $\mathbb{P}\left(d(\mathbf{X}) \geq t_{n-2}\right)$

$$\begin{aligned}\mathbb{P}(d(\mathbf{X}) \geq t_n) &\leq (1-\delta)^{-3(1+4)}\left((1-\delta)^{-3}\mathbb{P}\left(d(\mathbf{X}) \geq t_{n-3}\right)^4\right)^{4^2} \\ &= (1-\delta)^{-3(1+4+4^2)}\mathbb{P}\left(d(\mathbf{X}) \geq t_{n-3}\right)^{4^3},\end{aligned}$$

and if we would do this another $n-3$ times, we would find that

$$\mathbb{P}(d(\mathbf{X}) \geq t_n) \leq (1-\delta)^{-3(1+4+\ldots+4^{n-1})}\mathbb{P}\left(d(\mathbf{X}) \geq t_0\right)^{4^n}.$$

Since $t_0 = 3\eta$, (3.3.1) tells us that we can be certain that $\mathbb{P}(d(\mathbf{X}) \geq t_0) \leq \delta$. Furthermore

$$-3(1+4+\ldots+4^{n-1}) = -3\sum_{k=0}^{n-1} 4^k = -3\frac{1-4^n}{1-4} = 1-4^n,$$

So because $(1-\delta) \in \left(\frac{1}{2}, 1\right)$,

$$(1-\delta)^{-3(1+4+\ldots+4^{n-1})} = \frac{1}{(1-\delta)^{4^n-1}} \leq \frac{1}{(1-\delta)^{4^n}}.$$

Put this all together and we find that

$$\mathbb{P}(d(\mathbf{X}) \geq t_n) \leq \left(\frac{\delta}{(1-\delta)}\right)^{4^n} = \exp\left(4^n \log\frac{\delta}{1-\delta}\right).$$

To get to the inequality as given by (3.3.2) we take a look again at $t_{n+1} = (2^{n+1}+1)3\eta$, for some rewriting gives us

$$\frac{t_{n+1}+3\eta}{12\eta} = 2^n.$$

After squaring both sides we get

$$\left(\frac{t_{n+1}+3\eta}{12\eta}\right)^2 = 4^n,$$

so if we take t such that $t_n \leq t \leq t_{n+1}$ we find that

$$4^n \geq \left(\frac{t}{12\eta}\right)^2 = \frac{t^2}{144\eta^2}.$$

Because $\delta \in \left(0, \frac{1}{2}\right)$ it follows that $\log\frac{\delta}{1-\delta} \leq 0$, therefore

$$\mathbb{P}(d(\mathbf{X}) \geq t) \leq \mathbb{P}(d(\mathbf{X}) \geq t_n) \leq \exp\left(4^n \log\frac{\delta}{1-\delta}\right) \leq \exp\left(\frac{t^2}{144\eta^2}\log\frac{\delta}{1-\delta}\right).$$

This is true for every $t_n \leq t \leq t_{n+1}$, but we can do this for every $n \geq 0$, so since $t_n \geq t_0 = 3\eta$ and $t_n \to \infty$ as $n \to \infty$, we have proven that inequality (3.3.2) holds for every $t \geq 3\eta$. $\square$

Corollary 3.5.
Let $\mathbf{X}$ *be* $\mathcal{I}$*-Gaussian on a measurable Abelian group* G*, then there is an* $\varepsilon > 0$ *such that*

$$\mathbb{E}\left(e^{\varepsilon d(\boldsymbol{X})^2}\right) < \infty$$

Proof. According to Proposition 1.3, we have

$$\mathbb{E}\left(e^{\varepsilon d(\mathbf{X})^2}\right) = \int_0^\infty \mathbb{P}(||\mathbf{X}|| \geq t) 2\varepsilon t e^{\varepsilon t^2} dt.$$

As long as the function d is not always equal to zero (if it is, then $\mathbb{E}\left(e^{\varepsilon d(\mathbf{X})^2}\right) = 1 < \infty$), there will always exist a $\delta \in \left(0, \frac{1}{2}\right)$ and $\eta > 0$ such that

$$\mathbb{P}(d(\mathbf{X}) \geq \eta) \leq \delta.$$

Therefore, according to Theorem 3.4, we know that

$$\mathbb{P}(d(\mathbf{X}) \geq t) \leq e^{C_{\eta,\delta} t^2}$$

for $t \geq 3\eta$, with $C_{\eta,\delta} < 0$.

$$\begin{aligned}
\int_0^\infty \mathbb{P}(d(\mathbf{X}) \geq t) 2\varepsilon t e^{\varepsilon t^2} dt &= \int_0^{3\eta} \mathbb{P}(d(\mathbf{X}) \geq t) 2\varepsilon t e^{\varepsilon t^2} dt + \int_{3\eta}^\infty \mathbb{P}(d(\mathbf{X}) \geq t) 2\varepsilon t e^{\varepsilon t^2} dt \\
&\leq \int_0^{3\eta} 2\varepsilon t e^{\varepsilon t^2} dt + \int_{3\eta}^\infty 2\varepsilon t e^{(C_{\eta,\delta}+\varepsilon)t^2} dt \\
&= e^{\varepsilon t^2}\Big|_0^{3\eta} + \frac{\varepsilon e^{(C_{\eta,\delta}+\varepsilon)t^2}}{C_{\eta,\delta}+\varepsilon}\Big|_{3\eta}^{\infty}.
\end{aligned}$$

It is easy to see that the first term equals $e^{9\varepsilon\eta^2} - 1$ which is finite. Now if we take $\varepsilon < |C_{\eta,\delta}|$, then $C_{\eta,\delta} + \varepsilon < 0$, and we won't be bothered by that infinity

$$\frac{\varepsilon e^{(C_{\eta,\delta}+\varepsilon)t^2}}{C_{\eta,\delta}+\varepsilon}\Big|_{3\eta}^{\infty} = -\frac{\varepsilon e^{9(C_{\eta,\delta}+\varepsilon)\eta^2}}{C_{\eta,\delta}+\varepsilon},$$

which is again finite. Therefore

$$\mathbb{E}\left(e^{\varepsilon d(\mathbf{X})^2}\right) < \infty.$$

□

Knowing that $\mathbb{E}\left(e^{\varepsilon d(\mathbf{X})^2}\right) < \infty$ for an $\varepsilon > 0$, we can also conclude that the $\mathbb{E}\left(e^{t d(\mathbf{X})}\right) < \infty$ for $|t| < \varepsilon$, since $d(\mathbf{X})^2 \geq d(\mathbf{X})$ when $d(\mathbf{X}) \geq 1$. Therefore, with an argument similar as the one for the moment-generating functions (see "Expectation and Moments"), we can conclude that $\mathbb{E}(d(\mathbf{X})^n) < \infty$ for every n and an $\mathcal{I}$-Gaussian random variable has finite moments.

Bibliography

[1] Aliprantis, C.D., Burkinshaw, O. (1998). *Principles of Real Analysis* (3$^{\text{rd}}$ edition). London, UK: Academic Press

[2] Bauer, H. (1991). *Wahrscheinlichkeitstheorie*, Berlin, Germany: Walter de Gruyter

[3] Berge, C. (1971). *Principles of Combinatorics*, New York, New York: Academic Press

[4] Bobkov, G., Houdré, C. (1995). Open problem. *Stochastic Analysis Digest* **15**

[5] Bryc, W. (2005). *Normal Distribution characterizations with applications, Lecture notes in Statistics*, Cincinnati, Ohio: University of Cincinnati

[6] Cramér, H. (1946). *Mathematical Methods of Statistics*, Princeton, New Jersey: Princeton University Press

[7] Freitag, E., Busam, R. (2008). *Complex Analysis* (2$^{\text{nd}}$ edition). Berlin, Germany: Springer

[8] Kwapien, S., Pycia, M., Hachermayer, W. (1996). A proof of a conjecture of Bobkov and Houdre. *Electronic Communications in Probability*, **1**, 7-10

[9] Oleskiewicz, K. (1996). On certain characterization of normal distribution. *Statistics & Probability Letters*, 33, 277-280

[10] Ruszel, W.M. (2018). *Lecture notes on Advanced Probability TW 3560*, Delft, The Netherlands: Delft Institute of Applied Mathematics

[11] Tortrat, A. (1981). Lois stables dans un groupe. *Ann. Inst. H. Poincaré Probab. & Statist*, 17, 51-61

[12] Zygmund, A. (1947). A Remark on Characteristic Functions. *The Annals of Mathematical Statistics*, Vol. 18 (2), 272-276