Mathematics for Supply Chain

Msc Supply Chain & Purchasing

Guillaume Metzler
Institut de Communication (ICOM)
Université de Lyon, Université Lumière Lyon 2
Laboratoire ERIC UR 3083, Lyon, France
guillaume.metzler@univ-lyon2.fr

1 Course summary

In this first section, we provide a brief reminder of the concept studied in class during the first lecture which focus on the normal distribution.

1.1 Normal Distribution

This part review the definition of a gaussian (or normal) distribution and its properties. It is also the occasion to do some recalls on basic statistics and probabilities.

Density and Cumulative functions A random Variable $X \sim \mathcal{N}(\mu, \sigma^2)$, where μ is the mean value of the normal distribution and σ is the standard deviation (thus σ^2 is the variance) has a *probability density function* (*pdf* or density function in short) the following function f :

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(x-\mu)^2}{\sigma^2}\right).$$

The *continuous density function* (*cdf* or cumulative density function, associated to the density function, denoted by F is defined by :

$$F(t) = \mathbb{P}[X \leq t] = \int_{-\infty}^{t} f(x)dx = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(x-\mu)^2}{\sigma^2}\right) dx.$$

As you may notice, this continuous density function F gives the opportunity to compute probabilities compared to the density. It is also possible to represent this probability graphically by the area under the curve of the density (this the definition of the previous integral) as represented in Figure 1.

The continuous density function does not always an explicit expression and this is unfortunately the case for the normal distribution. The *cdf* and *pdf* are respectively represented in Figure 2.

Thus the probabilities $\mathbb{P}[X \leq t]$ are no more than computation of the F values F for a given real number t. For instance for $t = 1.38$, we have $F(1.38) = \mathbb{P}[X \leq 1.38] = 0.9162$.

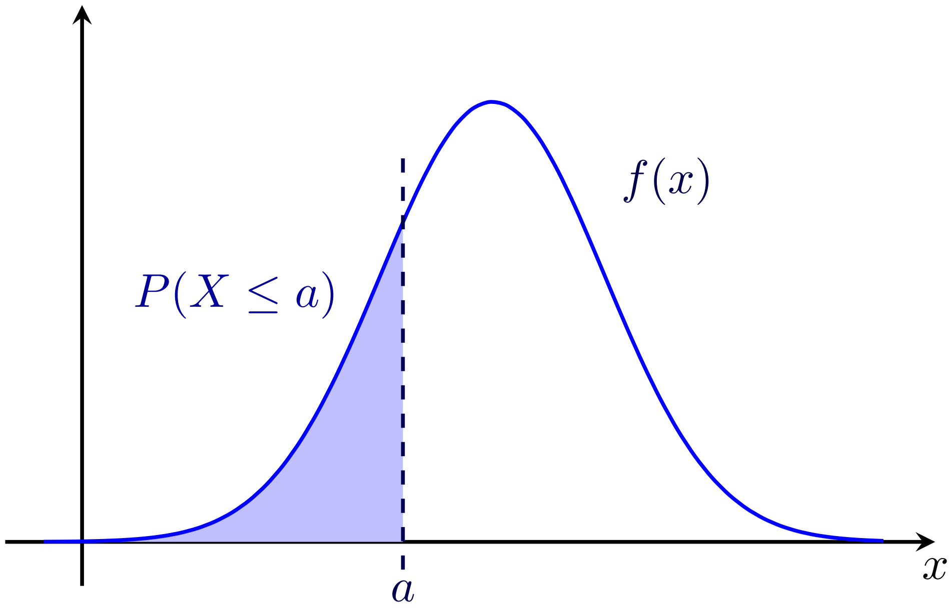

FIGURE 1 – Representation of the probability $\mathbb{P}[X \leq a]$ for a real number a using the density function where X is normally distributed.

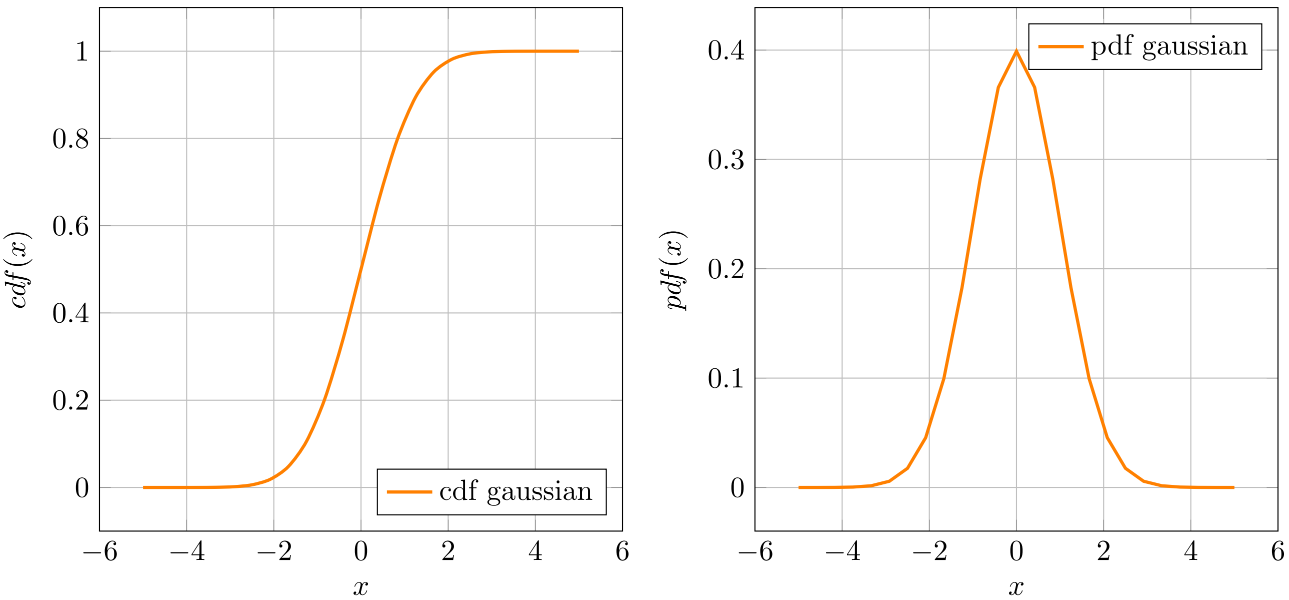

FIGURE 2 – Representation of the function F and f respectively for a random variable $X \sim \mathcal{N}(0,1)$ that follows a normal distribution with a mean equal to 0 and a variance equal to 1

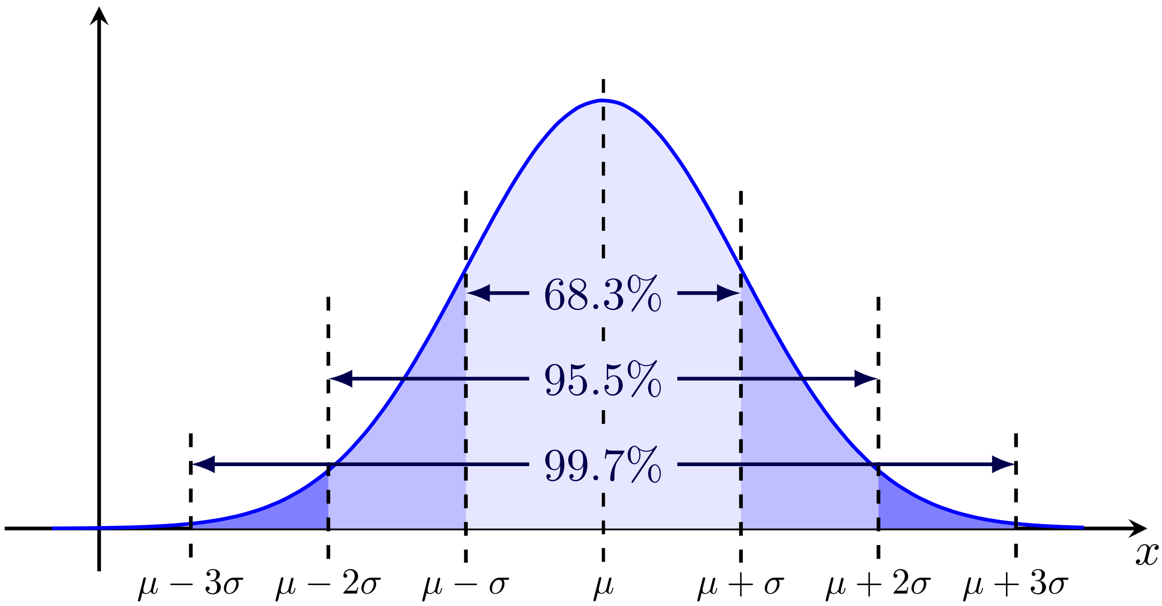

FIGURE 3 – Representation of the proportion of the values taken by a random variable following a normal distribution for different bandwidths defined by σ.

Some properties of the Normal Distribution A random variable X following a normal distribution with a mean μ and a variance σ^2 (*i.e.* with standard deviation σ) has the following properties :

- **Mean = Median = Mode** .
- **Symmetry :** for all real number t : $\mathbb{P}[X \leq \mu - t] = \mathbb{P}[X \geq \mu + t]$.

We also recall that for all real number t, and for all random variable X, we have :

$$\mathbb{P}[X \leq t] = 1 - \mathbb{P}[X \geq t].$$

If we come back to Figure 1, it means that the surface of blue area is equal to 1 minus the surface of the white area.
Also, for all real $t_1 < t_2$ we have :

$$\mathbb{P}[t_1 \leq X \leq t_2] = \mathbb{P}[X \leq t_2] - \mathbb{P}[X \leq t_1].$$

We finish by some important figures :

- 95% of the values taken by $X \sim \mathcal{N}(\mu, \sigma^2)$ fall in the rang $[\mu - 1.96\sigma, \mu + 1.96\sigma]$
- 99% of the values taken by $X \sim \mathcal{N}(\mu, \sigma^2)$ fall in the range $[\mu - 2.58\sigma, \mu + 2.58\sigma]$

All these values are summarized in Figure **??**

1.2 Computing Probabilities and Quantiles of a Normal Distribution

The process that we illustrate here is mainly used to make confidence estimation on a unknown parameter.

Normlization We assume that we have a random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ then the random variable Z defined by :

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

also follows a normal distribution but with $\mu = 0$ and $\sigma = 1$. We say that have *centered* and *reduce* our random variable X, and this process is mainly used when you want to estimate

$$\mathbb{P}[X \leq t] \quad \text{or} \quad \mathbb{P}[t_1 \leq X \leq t_2],$$

and you do not have access to an appropriate software but just a table.

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

Démonstration. For all real numbers a, we have the following properties on the l'**expectation** and on the **variance** of a random variable X

(i) $\mathbb{E}[X + a] = \mathbb{E}[X] + a$,
(ii) $\mathbb{E}[aX] = a\mathbb{E}[X]$,
(iii) $Var(a + X) = Var(X)$,
(iv) $Var(aX) = a^2 Var(X)$.

Thus, if $X \sim \mathcal{N}(\mu, \sigma^2)$, according to (i), the expectation of the random variable $X' = X - \mu$ is equal to 0, and its variance remains unchanged according to (iii) The expectation of the random variable $Z = \frac{X'}{\sigma} = \frac{X - \mu}{\sigma}$ is also equal to 0 according to (ii) and its variance is equal to 1 according to (iv). □

The way the different steps works on the distribution are illustrated on Figure 4.

Compute Probabilities In order to proceed to the Z normalization and to read the Z *table* and even for the use of Excel, it is necessary to have inequalities of the form

$$\mathbb{P}[Z \leq t]$$

to use them.

Exemple 1.1. *Let $X \sim \mathcal{N}(5, 2)$ et let us compute the probability that X takes values lower or equal than 1, i.e.* $\mathbb{P}[X \leq 1]$.

$$\begin{aligned}
& \quad \downarrow \textit{ we substract by the mean of } X \textit{ then we divide by its standard deviation on both sides of the inequality} \\
\mathbb{P}[X \leq 1] &= \mathbb{P}\left[\frac{X-5}{2} \leq \frac{1-5}{2}\right], \\
& \quad \downarrow \textit{ we simplify the right hand-side} \\
&= \mathbb{P}\left[\frac{X-5}{2} \leq -2\right], \\
& \quad \downarrow \textit{ we set } Z = \frac{X-5}{2} \textit{ and } Z \sim \mathcal{N}(0,1)
\end{aligned}$$

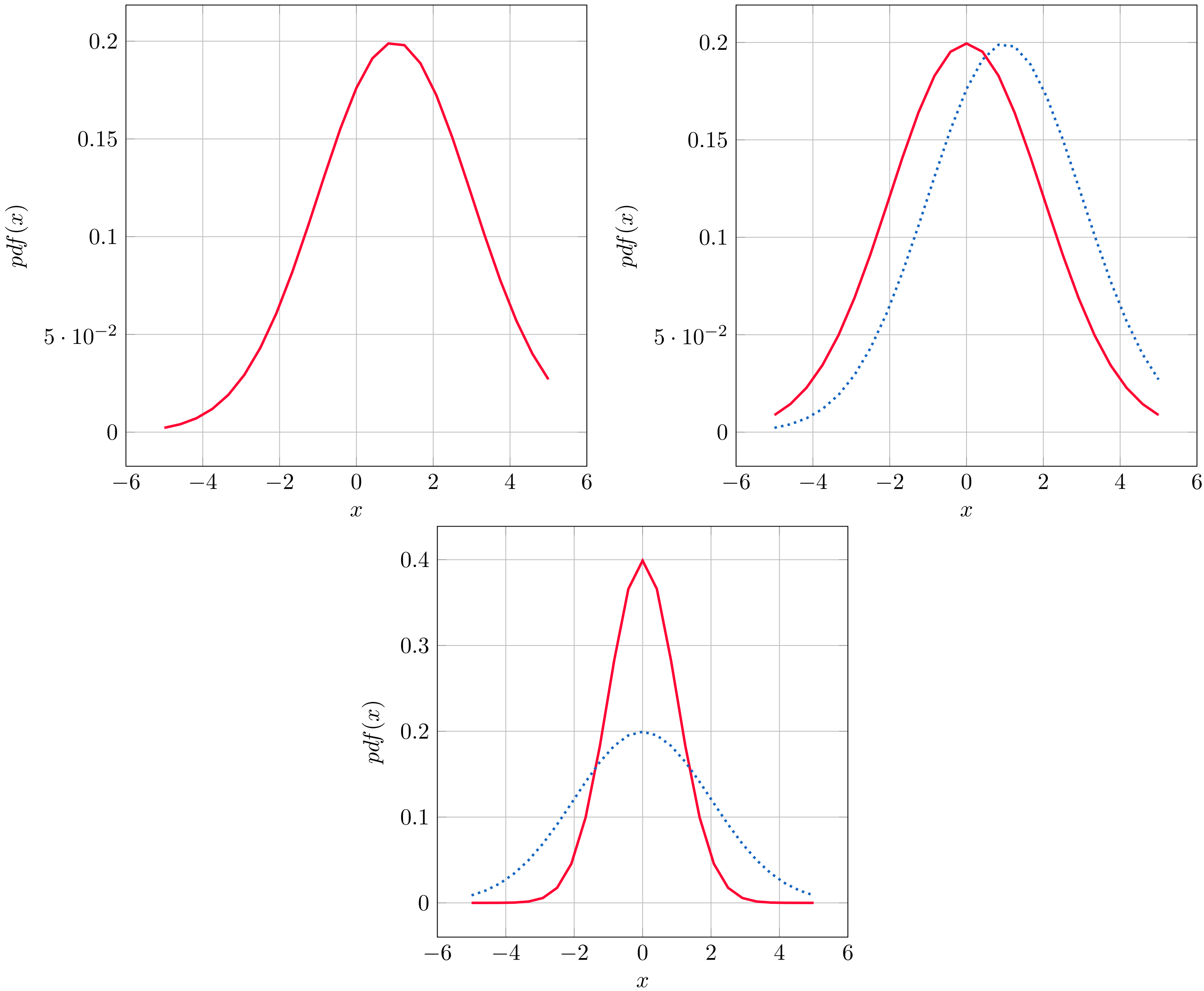

FIGURE 4 – Illustration (from the top left to the bottom) of the normalization steps The first figure, on the top left, shows the initial gaussian distribution The one on the top rights shows the transformation which consists in centering the distribution The one at the bottom illustrates the impact of the reduction of this normal distribution, in order to achieve a variance equal to 1. On both figures, the dotted lines represent the distribution at the previous step.

$$= \mathbb{P}[Z \leq -2],$$

$\downarrow$ *se use the symmetry*

$$= \mathbb{P}[Z \geq 2],$$

$\downarrow$ *we use the fact that* $\mathbb{P}[Z \leq t] = 1 - \mathbb{P}[Z \geq t]$

$$= 1 - \mathbb{P}[Z \leq 2],$$

$\downarrow$ *we can look at the value on the Z-table or using Excel*

$$= 1 - 0.977,$$

$$= 0.023.$$

With Excel

With Excel, we can compute the quantiles $F^{-1}(p)$, where p is probability, associated to a gaussian distribution using the following formula :

Version	Function
ENGLISH	NORM.DIST(t, μ, σ, CUMULATIVE)
FRENCH	LOI.NORMALE.N(t ; μ ; σ ; CUMULATIVE)

where μ and σ are respectively the mean and the standard deviation of the normal distribution. The last parameter is called Cumulative, if :

- **TRUE** : it computes $\mathbb{P}[X \leq t] = F(t)$, otherwise
- **FALSE** : it computes $f(t)$ the value of the density for the given t.

Compute Quantiles It sometimes interesting, mostly when it comes to build confidence intervals for estimation or also for hypothesis testing, to be able to find a so called quantile.
The reader can see it as the the reverse of the previous process, *i.e.*, previously, we aim to compute $\alpha = \mathbb{P}[X \leq t] = F(t)$ for a given t. But this time, we want to find the value of t, given $\alpha \in [0, 1]$, such that $F(t) = \alpha$, *i.e.*, $t = F^{-1}(\alpha)$.

For instance if we take $\alpha = 0.5$, then $F^{-1}(\alpha)$ is the just the median, *i.e.* the quantile of order 0.5. More generally, the value t_α such that $\alpha = \mathbb{P}[X \leq t_\alpha]$ is called the quantile of order α.

There is now way to have directly access to this value and we need a software fort that.

With Excel

With Excel, we can compute the quantiles $F^{-1}(p)$, where p is probability, associated to a gaussian distribution using the following formula :

Version	Function
ENGLISH	NORM.INV(p, μ, σ)
FRENCH	LOI.NORMALE.INVERSE.N(p ; μ ; σ)

where μ and σ are respectively the mean and the standard deviation of the normal distribution and $p \in [0, 1]$ is the level of the quantile.

2 Exercises

Exercice 2.1. *Ages in a city are considered to be distributed according to a normal distribution* $\mathcal{N}\left(\mu, \sigma^2\right)$*, with the values of* μ *and* σ *depending on the city. An individual is considered young if he is under* 25 *and old if he's over* 50*. Finally, an individual whose age is between* 30 *and* 40 *will be considered to be in "golden age".*

We assume that the age of the population in Strasbourg follows a normal distribution with parameters $\mu = 33$ *and* $\sigma = 20$*, whereas in Lyon, the population follows a normal distribution with parameters* $\mu = 35$ *and* $\sigma = 22$*.*

a) What are the limitations of this model ?
b) In which city can we find the most young people ?
c) Conversely, in which city can we find the most old people ?
d) Finally, in which city can we find the most people in their prime ?

Exercice 2.2 (Probabilities and Costs)**.**
We want to study the distribution of supply times (expressed in hours) between a supplier and a set of stores for a particular brand for the coming year.
Studies carried out the previous year showed that these times, are distributed according to a normal law of mean $\mu = 1$ *(and standard deviation* $\sigma = 0.1$*.*

What is the probability that the supply time is :

1. *lower than one hour*
2. *between* 0.95 *and* 1 *hour*
3. *between* 1 *and* 1.05 *hour*
4. *lower than* 0.95 *or greater than* 1.05*.*

It is assumed that the supply cost is time dependent and represented by a random variable :

- *the cost is equal to* $1,000\$$ *when the the supply time is lower than* 0.95
- *the cost is equal to* $1,200\$$ *when the the supply time is between than* 0.95 *and* 1
- *the cost is equal to* $1,400\$$ *when the the supply time is between than* 1 *and* 1.05
- *the cost is equal to* $1,500\$$ *when the the supply time is greater than* 1.05

Determine the mean average cost using the previous questions' results.

Exercice 2.3 (Study of an Assembly line)**.**
We consider the production of $85g$ *cans. The process includes : filling in the cans, sealing, weighing and disposal if the weight is less than* $85g$ *(production waste), sorting, labeling. Even if the automatons are of very good quality, it is impossible to obtain cans weighing exactly* $85g$*. There is a standard deviation in packing the cans related to the technical processes, which follows a normal distribution of the standard deviation equal to* $0.8g$*.*

1. *We aim to determinate the mean value of packing the cans that has to be used so that the waste related to the weight of package is less than* 2%*.*

We are now interested in the production of $100g$ *packages. Once the mixture has been prepared, it is packaged by the automatons. The process includes : packing, sealing, weighing and disposal if the weight is less than* $98.5g$ *or bigger than* $101.5g$*, sorting, labeling. The automatons are the same as those used for the* $85g$ *packages. The standard deviation in packing, related to the technical processes, follows, also for the* $100g$ *packages, a normal distribution of the standard error equal to* $0.8g$*.*

2. *Knowing that the automatons are set to a mean* $\mu = 100g$*, we are interested in the percentage of packages that are then discarded as waste, statistically speaking.*
3. *Would it be appropriate to set the machines to a mean* $\mu = 100.5$ *?*

Exercice 2.4 (Probabilities)**.** *You were told that the amount of time lapsed between consecutive commands on a website followed a normal distribution with a mean of* 15 *seconds. You were also told that the probability that the time lapsed between two consecutive commands to fall between* 16 *to* 17 *seconds was* 13%*. The probability that the time lapsed between two consecutive commands would fall below* 13 *seconds was* 7%*.*

What is the probability that the time lapsed between two consecutive commands will be between 13 *and* 16 *seconds ?*

Exercice 2.5. *It is assumed that the average distance covered by a delivery driver during a delivery day follows a Gaussian distribution. During a first period of measure, we find that* 10% *of all journeys are longer than* 250*km and exactly* 25% *of the journey are less longer than* 150*km.*

Compute the average distance covered by the delivery driver, and the standard deviation of this distance.