Monte Carlo Project 1 : "Negative Weighted Mixture"

Samuel Elbaz, Alexandre Zenou,, Sacha Assouly

October 2024

Contents

Exercise

Consider a random variable of a negative weighted mixture X following a density law proportional to

$$\forall x \in \mathbb{R}, \quad f(x) \propto f_1(x) - a f_2(x)$$

with $f_1(x) = \mathcal{N}(x; \mu_1, \sigma_1^2)$ and $f_2(x) = \mathcal{N}(x; \mu_2, \sigma_2^2)$ density laws of two normal distributions; and $a > 0$. The objective of this exercise is to understand the behavior of the random variable and to compute statistics.

1 Part I : Definition

1.1 Question 1

Recall the conditions for a function f to be a probability density. Considering the tail behavior of f, derive necessary conditions on (σ_1^2, σ_2^2) for f to be a density.

$f : \mathbb{R} \to \mathbb{R}$ is a probability density with respect to λ the Lebesgue measure if it satisfies :

1. $$f \geq 0 \ \lambda\text{-almost everywhere}$$

2. $$\int_{\mathbb{R}} f \, d\lambda = 1$$

To search for necessary conditions such that f is a probability density, we start by checking its positivity. Let us consider the inequality :

$$f_1(x) - a f_2(x) \geq 0, \quad \forall x \in \mathbb{R}$$

where f_1 and f_2 are the probability density functions of two normal distributions:

$$f_1(x) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right), \quad \forall x \in \mathbb{R}$$
$$f_2(x) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right), \quad \forall x \in \mathbb{R}$$

Here, σ_1^2 and σ_2^2 are the variances of the respective normal distributions, and since variances are always positive, we have $\sigma_1 > 0$ and $\sigma_2 > 0$.

This inequality simplifies to :

$$\exp\left(\frac{(x-\mu_2)^2}{2\sigma_2^2} - \frac{(x-\mu_1)^2}{2\sigma_1^2}\right) \geq a\frac{\sigma_1}{\sigma_2}, \quad \forall x \in \mathbb{R}.$$

The exponent is a second-degree polynomial:

$$
\begin{aligned}
P(x) &= \frac{(x-\mu_2)^2}{2\sigma_2^2} - \frac{(x-\mu_1)^2}{2\sigma_1^2} \\
&= \left(\frac{1}{2\sigma_2^2} - \frac{1}{2\sigma_1^2}\right)x^2 + \left(\frac{-\mu_2}{\sigma_2^2} + \frac{\mu_1}{\sigma_1^2}\right)x + \left(\frac{\mu_2^2}{2\sigma_2^2} - \frac{\mu_1^2}{2\sigma_1^2}\right)
\end{aligned}
$$

For the exponential to remain greater than $a\frac{\sigma_2}{\sigma_1}$, the highest degree coefficient must be strictly positive:

$$\frac{1}{2\sigma_2^2} - \frac{1}{2\sigma_1^2} > 0$$

For this expression to be positive, we must have :

$$\sigma_1^2 > \sigma_2^2 > 0$$

Thus, the variance of the first distribution must be strictly greater than the variance of the second distribution to satisfy the positivity condition of f.

If this coefficient were negative (i.e., $\sigma_1^2 < \sigma_2^2$), the exponential would decay as x increases, tending to 0 as $x \to \pm\infty$, and would eventually become smaller than $a \cdot \frac{\sigma_1}{\sigma_2}$, violating the inequality. Therefore, $\sigma_1^2 > \sigma_2^2$ is necessary for the expression to remain positive and for the inequality to hold for all $x \in \mathbb{R}$.

Furthermore, since f_1, f_2 belong to $L^1(\mathbb{R})$, which is a vector space, we have that f is in $L^1(\mathbb{R})$ and integrable as a linear combination of f_1 and f_2. We thus have our necessary conditions on (σ_1^2, σ_2^2) to check in order to respect the integrability condition.

1.2 Question 2

For given parameters $\theta_1 = (\mu_1, \sigma_1^2)$ and $\theta_2 = (\mu_2, \sigma_2^2)$, determine a bound a^* on a for $f(x) \propto f_1(x) - af_2(x)$ to be a well-defined density. Provide its normalization constant.

Let us take back the inequality from the previous question :

$$\exp\left(\frac{(x-\mu_2)^2}{2\sigma_2^2} - \frac{(x-\mu_1)^2}{2\sigma_1^2}\right) \geq a\frac{\sigma_1}{\sigma_2}, \quad \forall x \in \mathbb{R}$$

From this expression, we can major a :

$$a \leq \frac{\sigma_2}{\sigma_1} \cdot \exp\left(\frac{(x-\mu_2)^2}{2\sigma_2^2} - \frac{(x-\mu_1)^2}{2\sigma_1^2}\right), \quad \forall x \in \mathbb{R}$$

We can now conclude that a is bounded by the infimum of the expression on the left-hand side of the inequality for all $x \in \mathbb{R}$. This gives us:

$$a \leq \inf_{x\in\mathbb{R}} \left(\frac{\sigma_2}{\sigma_1} \cdot \exp\left(\frac{(x-\mu_2)^2}{2\sigma_2^2} - \frac{(x-\mu_1)^2}{2\sigma_1^2}\right)\right) \tag{1}$$

Since exp is strictly increasing on $\mathbb{R}$, this comes down to studying the problem:

$$\inf_{x\in\mathbb{R}} P(x)$$

with

$$P(x) = \left(\frac{1}{2\sigma_2^2} - \frac{1}{2\sigma_1^2}\right)x^2 + \left(\frac{-\mu_2}{\sigma_2^2} + \frac{\mu_1}{\sigma_1^2}\right)x + \left(\frac{\mu_2^2}{2\sigma_2^2} - \frac{\mu_1^2}{2\sigma_1^2}\right)$$

Let's first remark that this problem has a unique solution as P is strongly convex on $\mathbb{R}$ as an even-degree polynomial. To find the minimum of the polynomial without deriving, we use the summit formula:

$$x_{\min} = -\frac{b}{2a}$$

where:

- a is the second degree coefficient,
- b is the first degree coefficient.

In our case, we have:

$$a = \left(\frac{1}{2\sigma_2^2} - \frac{1}{2\sigma_1^2}\right), \quad b = \left(\frac{-\mu_2}{\sigma_2^2} + \frac{\mu_1}{\sigma_1^2}\right).$$

The calculation for the minimum becomes:

$$x_{\min} = \frac{-\left(\frac{-\mu_2}{\sigma_2^2} + \frac{\mu_1}{\sigma_1^2}\right)}{2\left(\frac{1}{2\sigma_2^2} - \frac{1}{2\sigma_1^2}\right)}.$$

Simplifying this we get:

$$x_{\min} = \frac{\mu_2\sigma_1^2 - \mu_1\sigma_2^2}{\sigma_1^2 - \sigma_2^2}.$$

Noting a^* as the infimum of equation (1), the computation becomes:

$$a^* = \frac{\sigma_2}{\sigma_1} \cdot \exp\left(\frac{(x_{\min} - \mu_2)^2}{2\sigma_2^2} - \frac{(x_{\min} - \mu_1)^2}{2\sigma_1^2}\right)$$

Simplifying, we get:

$$a^* = \frac{\sigma_2}{\sigma_1} \cdot \exp\left(\frac{\sigma_2^2(\mu_2 - \mu_1)^2}{2(\sigma_1^2 - \sigma_2^2)^2} - \frac{\sigma_1^2(\mu_2 - \mu_1)^2}{2(\sigma_1^2 - \sigma_2^2)^2}\right)$$

And finally:

$$a^* = \frac{\sigma_2}{\sigma_1} \cdot \exp\left(\frac{-(\mu_2 - \mu_1)^2}{2(\sigma_1^2 - \sigma_2^2)}\right)$$

This relation expresses the upper bound a^* that we are seeking, which is the maximum value that a can take in order for the inequality to hold. Lets remark that $a^* < 1$ since :

- $\frac{\sigma_2}{\sigma_1} < 1$
- the exponant is negative

Knowing the upper bound of a we can now deduce the value of the normalizing constant in $f \propto f_1 - af_2$. To compute the normalizing constant $C \in \mathbb{R}_+^*$, we have:

$$C \int_{-\infty}^{\infty} (f_1(x) - af_2(x))\, dx = 1$$

Since both f_1 and f_2 are probability densities, their integrals over $\mathbb{R}$ are 1:

$$C(1 - a) = 1$$

Since $a \in]0; a^*] \subset]0; 1[$, the normalizing constant is well defined and is :

$$C := \frac{1}{1 - a}$$

1.3 Question 3

Create an R *function* `f(a, mu_1, mu_2, s_1, s_2, x)` *that produces the pdf of f as a function of x and of the parameters $a, \mu_1, \mu_2, \sigma_1, \sigma_2$. Plot on the same graph the pdf of f for different values of a, especially for $a = a^*$. In another graph, do the same for different values of σ_2, for σ_1 fixed. Examine the impact of a and σ_2 on the shape of f.*

We first start by coding the `R` function `f` that produces the probability distribution function of f:

```
```{r}
#We define the upper bound function of a
compute_a_star <- function(mu_1, mu_2, s_1, s_2){
  return((s_2 / s_1) * exp(-1 / 2 * (mu_1 - mu_2)^2 / ((s_1)^2 - (s_2)^2)))
}

# We define the density function f for the weighted mixture
f <- function(a, mu_1, mu_2, s_1, s_2, x) {
  #We compute a_start to verify a is in the interval
  a_star <- compute_a_star(mu_1,mu_2,s_1,s_2)
  if(a > a_star){a <- a_star }
  # We use dnorm to calculate the normal densities f1 and f2
  f1 <- dnorm(x, mean = mu_1, sd = s_1)
  f2 <- dnorm(x, mean = mu_2, sd = s_2)

  # We return the weighted mixture of f1 and f2
  return(1/(1 - a) * (f1 - a * f2) )
}
```

Figure 1: R function `f(a, mu_1, mu_2, s_1, s_2, x)` to generate the pdf of f

We set the parameters as follows:

- $\mu_1 = 0$
- $\mu_2 = 1$
- $\sigma_1 = 2$
- $\sigma_2 = 1$

Since a lies within the interval $(0; a^*]$, we vary a to graphically observe the evolution of f for different values of a :

We graphically observe that:

- When $a = a^*/100$, the density corresponds almost entirely to f_1, which follows a normal distribution $\mathcal{N}(0, 4)$. This explains why the curve is centered around 0 and more spread out.
- When $a = a^*/2$, the peak shifts slightly to the left due to the influence of $-f_2$, which follows a normal distribution $\mathcal{N}(-1, 1)$. The distribution becomes less symmetric and starts showing a dip near the mean of f_2.
- When $a = a^*$, the effect of f_2 is at its maximum. The peak shifts further to the left, and the curve shows a more pronounced dip around the mean of f_2. This reflects the increasing dominance of f_2, leading to a left-skewed distribution with a lower peak.

Thus, as a grows, the influence of $-f_2$ becomes stronger which shifts the distribution leftwards.
```

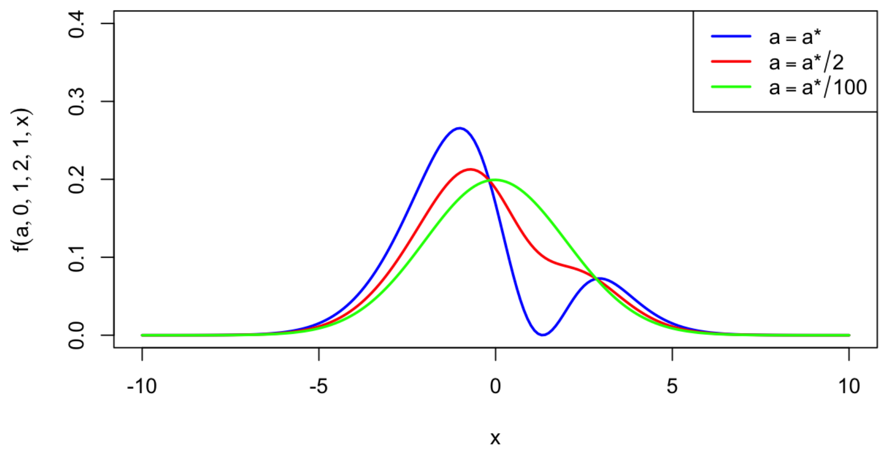

Figure 2: Density function f with varying values of a

We now look at the graphic behavior of f when varying σ_2 in $(0;\sigma_1)$ with σ_1 fixed. We keep the same parameters as previously and set $a = a^*$. As we just saw, this should let us see the effect of f_2 since a^* maximizes the influence of f_2.

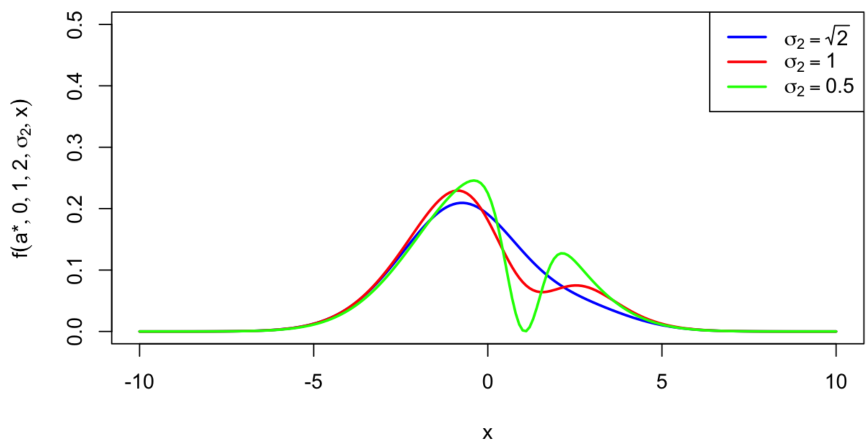

Figure 3: Density function f with varying values of σ_2 for fixed σ_1

We graphically observe that:

- When $\sigma_2 = \sqrt{2}$, f_1 and f_2 both have high variances. Even though their means are distinct, they are close enough to have their curves overlap to generate a single peak the density. f is then of high variance which explains why its peak is not sharp.
- When $\sigma_2 = 1$, the density start concentrating around the peaks of f_1 and f_2. Thus the peak of the distribution is sharper and a lump appears on the right tail. This corresponds to the influence of f_1.
- When $\sigma_2 = 0.5$, the density is concentrated around the peaks of f_1 (on the left) and f_2 (on the right). At this point, despite the closeness of μ_1 and μ_2, the variance of f_2 is so small that its density barely overlaps the density of f_1. Thus the two distinct peaks of f's curve.

Thus, as σ_2 decreases, the influence of f_2 becomes more localized around its mean. Since f_1 and f_2 have distinct means, this progressively creates two peaks on the curve of the pdf.

2 Part II : Inverse C.D.F Random Variable

We now set the parameters as follows: $\mu_1 = 0$, $\mu_2 = 1$, $\sigma_1^2 = 9$, $\sigma_2^2 = 1$. Using these values, we compute:

$$a^* = \frac{1}{3} \cdot \exp\left(\frac{-1}{4}\right) \approx 0.26$$

$a := 0.2$ is in $[0; a^*]$, so theses values are compatible with the constraints.

2.1 Question 4

:*Show that the cumulative density function associated with f is available in closed form. Create an R function `F(a, mu_1, mu_2, s_1, s_2, x)` that produces the cdf of f as a function of x and of the parameters a, μ_1, μ_2, σ_1, σ_2. Construct an algorithm that returns the value of the inverse function method as a function of $u \in (0,1)$, of the parameters a, μ_1, μ_2, σ_1, σ_2, and of an approximation precision ϵ. Deduce an algorithm that implements the inverse function method for the generation of random variables from F.*

Lets recall that f is of the form:

$$f(x) = \frac{1}{1-a}\left(f_1(x) - a \cdot f_2(x)\right), \quad \forall x \in \mathbb{R}$$

Thus, its cumulative distribution function F is:

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \frac{1}{1-a}\left(\int_{-\infty}^{x} f_1(t)\,dt - a\int_{-\infty}^{x} f_2(t)\,dt\right) \quad \forall x \in \mathbb{R}$$

Since $f_1 \sim \mathcal{N}(0,9)$ and $f_2 \sim \mathcal{N}(1,1)$, we have that:

$$F(x) = \frac{1}{1-a}\left(\Phi\left(\frac{x}{3}\right) - a \cdot \Phi(x-1)\right), \quad \forall x \in \mathbb{R}$$

where Φ is the CDF of a $\mathcal{N}(0,1)$.

Finally, substituting the values $a = 0.2$, we get the following closed-form expression for the F:

$$F(x) = 1.25 \cdot \Phi\left(\frac{x}{3}\right) - 0.25 \cdot \Phi(x-1), \quad \forall x \in \mathbb{R}$$

Here is F defined in R:

````
```{r}
# We implement the dichotomy method to invert F
bisection <- function(u, a, mu_1, mu_2, s_1, s_2, lower_bound = -50, upper_bound = 30, tol = 1e-7) {
  # We initialize the lower and upper bounds
  lower <- lower_bound
  upper <- upper_bound

  # We iteratively narrow down the interval until the desired precision is reached
  while ((upper - lower) > tol) {
    mid <- (lower + upper) / 2
    F_mid <- F(a, mu_1, mu_2, s_1, s_2, mid)

    # If F(mid) is less than u, the solution lies in the upper half
    if (F_mid < u) {
      lower <- mid
    } else {
      # Otherwise, the solution lies in the lower half
      upper <- mid
    }
  }

  # We return the midpoint as the approximate solution
  return((lower + upper) / 2)
}
```
````

Figure 4: F cdf definition in R

And here is it's graph:

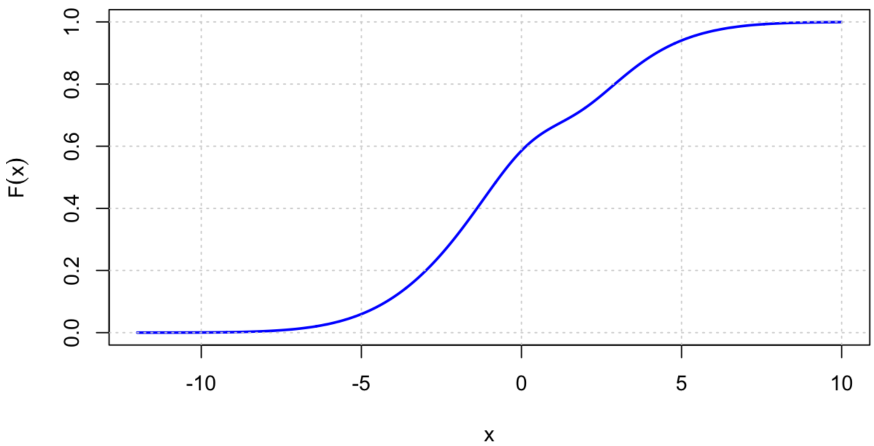

Figure 5: Plot of F

We observe two inflexion points around 0 which correspond f's double peaks.

We now consider u as the realization of a uniform random variable on $[0, 1]$. Given F, we seek $x \in \mathbb{R}$ such that:

$$F(x) = u$$

F is the cdf of a continuous law, as linear combination of the cumulative distribution functions of normal distributions, and is strictly increasing, so is invertible by bijection theorem for real valued functions, and x is the solution of:

$$x = F^{-1}(u)$$

However, we do not have the expression for the inverse of F. Therefore, we will apply the bisection algorithm to solve $F(x) - u = 0$. Here is its implementation in R:

```{r}
# We implement the dichotomy method to invert F
bisection <- function(u, a, mu_1, mu_2, s_1, s_2, lower_bound = -50, upper_bound = 30, tol = 1e-7) {
  # We initialize the lower and upper bounds
  lower <- lower_bound
  upper <- upper_bound

  # We iteratively narrow down the interval until the desired precision is reached
  while ((upper - lower) > tol) {
    mid <- (lower + upper) / 2
    F_mid <- F(a, mu_1, mu_2, s_1, s_2, mid)

    # If F(mid) is less than u, the solution lies in the upper half
    if (F_mid < u) {
      lower <- mid
    } else {
      # Otherwise, the solution lies in the lower half
      upper <- mid
    }
  }

  # We return the midpoint as the approximate solution
  return((lower + upper) / 2)
}
```

Figure 6: R bisection function to find $F^{-1}(u)$

Remark that the upper and lower bound are set to -12 and 10 in order for $F^{-1}(u)$ to be in [lower_bound;upper_bound] with a probability close to 1. We must have:

- $\mathbb{P}(F^{-1}(u) < \text{lower\_bound}) \approx 0 \Leftrightarrow u < F(\text{lower\_bound}) \approx 0$
- $1 - \mathbb{P}(F^{-1}(u) < \text{upper\_bound}) \approx 0 \Leftrightarrow 1 - (u < F(\text{upper\_bound})) \approx 0$

The computation using the R code of F confirms the bounds are fine:

```
#We check that the F is small (resp big) enough at lower_bound (resp upper_bound)
F(a, mu_1, mu_2, s_1, s_2,lower_bound)
1-F(a, mu_1, mu_2, s_1, s_2,upper_bound)

```

[1] 0.00003958905
[1] 0.0005363254
```

**Figure 7: Numerical check of the upper and lower bounds**

Furthermore the method converges linearly at rate $\alpha = \frac{1}{2}$. Indeed, noting $I_n$ to size of the interval at time $n \in \mathbb{N}$, we have that:

$$\begin{cases} I_0 = \text{upper\_bound} - \text{lower\_bound} \\ I_{n+1} = \frac{I_n}{2}. \end{cases}$$

$(I_n)_{n\in\mathbb{N}}$ is a geometrical sequence and thus:

$$I_n = \frac{\text{upper\_bound} - \text{lower\_bound}}{2^n}$$

By setting the tolerance to $\epsilon = 10{-}7$, we converge in:

$$n = \log_2(\text{upper\_bound} - \text{lower\_bound}) - \log_2(\epsilon) \approx 28$$

Repeating the process for $(u_1, \ldots, u_N)$ uniform law realizations, for $N \in \mathbb{N}$, we will obtain $(x_1, \ldots, x_N)$ solutions by the bisection algorithm that we will give us an approximation of $F^{-1}$. Since $F^{-1}(U)$ follows the law of $F$ for $U \sim$, $(x_1, \ldots, x_N)$ can be seen as $N$ realizations that follow $F$'s law. We therefore have a method to generate random variables from $F$.
```

2.2 Question 5

Write an R function ***inv_cdf(n)*** *that generates n samples from f using the inverse function method. Generate $n = 10000$ samples, and graphically check that* ***inv_cdf()*** *is correct.*

Following the path previously mentionned, we implement the function `inv_cdf`:

```
```{r}
# Define the function to generate n random variables
inv_cdf <- function(n, a, mu_1, mu_2, s_1, s_2) {
  u <- runif(n)  # Draw n uniform random variables
  sapply(u, function(x) bisection(x, a, mu_1, mu_2, s_1, s_2))
}
```
```

Figure 8: `inv_cdf` Generates n realizations of law F

From this function we may plot the invert function approximation of F^{-1}:

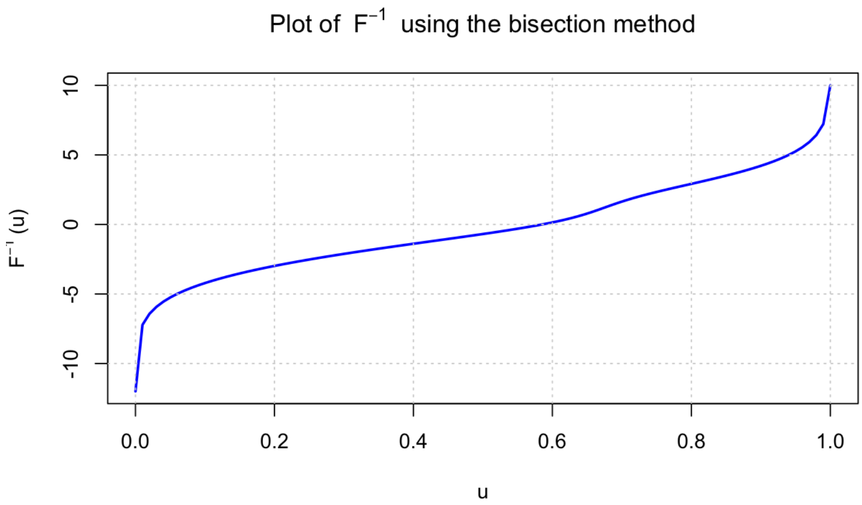

Figure 9: Plot of F^{-1} approximation generated with 10^4 samples

Finally we can graphically check that **inv cdf** is correct by comparing the repartition of the $(x_i)_i$ with the pdf f:

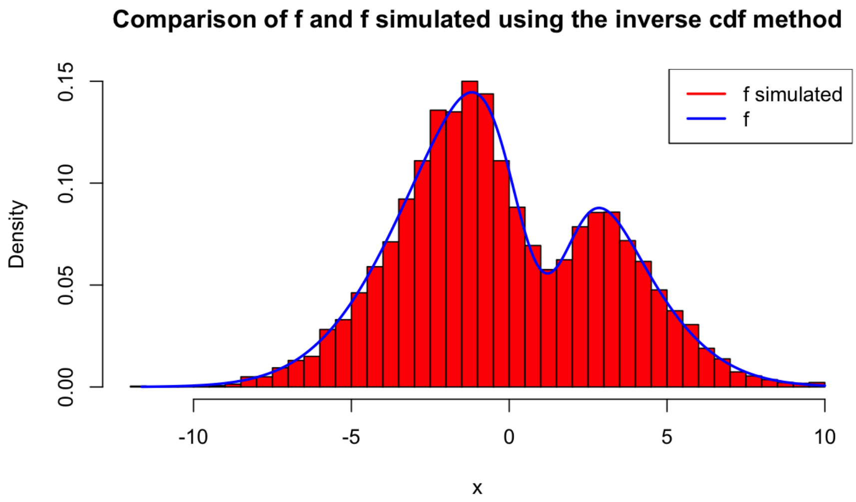

Figure 10: Comparison of f and f simulated

We observe that f follows the f's simulated histogram generated with our function `inv_cdf`. We graphically conclude that our function is correct.

3 Part III : Accept-Reject method

3.1 Question 6

Describe a method to simulate under f using the accept-reject algorithm and give the expression theoretical acceptance rate according to the parameters of f.

For the **Accept-Reject Method**, we first recall the very first constraints on what is called the *target density* : f, and the *instrumental dentisty* : g. Before stating the two conditions, we recall that we still consider the Lebesgue Measure λ on the measurable space $(\mathbb{R}; \mathcal{B}(\mathbb{R}))$. Then we have the following assumptions :

1. f and g have compatible supports :

$$\{f > 0\} \subseteq \{g > 0\}$$

[1]

2. There exists an $1 \leq M < \infty$ such that:

$$M \geq \tfrac{f}{g} \quad \text{a-e}$$

The method consists in finding the best g, such that it exists some $M > 0$ which gives $Mg(x) \geq f(x)$ a-e, then we generate uniform values living in the good set. More precisely, we take a proposal distribution g that is similar in shape to f, and easy to sample from (such as a normal distribution). M is chosen to ensure that g envelopes f. Then :

1. Draw a candidate $Y \sim g$.
2. Draw a uniform random variable $U \sim \mathcal{U}([0,1])$.
3. Accept Y if $U \leq \frac{f(Y)}{M \cdot g(Y)}$, otherwise reject and repeat the process until the condition holds.

The closer g is to f, the smaller M can be, and the higher the acceptance rate is. For optimal efficiency, g should be chosen to match the tails and peak of f as closely as possible.

In our case, we will have to choose a relevant g, which is not so complicated as we recall that our X follows the density f, which is given by:

$$f(x) = \frac{1}{1-a} \left(f_1(x) - a \cdot f_2(x) \right), \quad \forall x \in \mathbb{R}.$$

Thus, we can notice that $f_1(x) > f_1(x) - a \cdot f_2(x) > 0$ almost everywhere on $\mathbb{R}$. A good instrumental density could be this f_1. Define g as f_1 and f stays the same as in the exercise. Then, we want to find M such that $f_1 \cdot M \geq f$ almost everywhere. To this end, we study the following quantity:

$$\sup_{x \in \mathbb{R}} \frac{f(x)}{g(x)}.$$

From the previous assumptions, we know for sure that this term is well defined, indeed : if $f > 0$ then $g > 0$ a-e, so $\frac{f}{g}$ exists for a-e x in $\{f > 0\}$. In our case, $f > 0$ a-e, as the *sup* is taken over all $\mathbb{R}$, the negligeable values such that g vanishes does not bother us at all. Hence:

$$\begin{aligned}\frac{f(x)}{g(x)} &= \frac{\frac{1}{1-a}\left(f_1(x) - a \cdot f_2(x) \right)}{f_1(x)} \\ &= \frac{1}{1-a} \left(1 - a \frac{f_2(x)}{f_1(x)} \right).\end{aligned}$$

[1]For any h measurable over $(\mathbb{R}; \mathcal{B}(\mathbb{R}))$: $\{h > 0\} = \{x \in \mathbb{R}, h(x) > 0\}$

Expanding this ratio, we have

$$\frac{f_2(x)}{f_1(x)} = \frac{\sigma_1}{\sigma_2} \exp\left(-\frac{(x-\mu_2)^2}{2\sigma_2^2} + \frac{(x-\mu_1)^2}{2\sigma_1^2}\right).$$

Since $\sigma_2 < \sigma_1$, the term $-\frac{1}{2\sigma_2^2} + \frac{1}{2\sigma_1^2}$ in front of x^2 is negative. Therefore, as $x \to \pm\infty$, the exponential term tends to zero:

$$\lim_{|x|\to+\infty} \frac{f_2(x)}{f_1(x)} = 0.$$

Thus, we easily conclude that

$$\sup_{x\in\mathbb{R}} \frac{f(x)}{g(x)} = \frac{1}{1-a}.$$

We define M as the above quantity. Since a is in $]0, a^*]$ with $a* < 1$ by construction, it follows that $1 - a < 1$. Consequently, $M := \frac{1}{1-a} > 1$, as it is expected from the assumptions.

Remark: For every f and g are density functions, we necessarily have $M \geq 1$. Indeed, if $f(x) \leq Mg(x)$ almost everywhere, then

$$\int_{\mathbb{R}} f(x)\, dx \leq M \int_{\mathbb{R}} g(x)\, dx.$$

Since both f and g integrate to 1 (as they are densities), it follows that

$$1 \leq M$$

Alternatively, M can also be obtained through the acceptance-rejection criterion. For independent sequences $(U_n)_{n\geq1} \sim \mathcal{U}([0,1])$ and $(Y_n)_{n\geq1} \sim g$, we define

$$T := \inf\{n \geq 1 : U_n \leq \alpha(Y_n)\},$$

where $\alpha(Y_n) := \frac{f(Y_n)}{Mg(Y_n)}$. This construction ensures that Y_T follows the target density f, which implies that M must be large enough to satisfy $uMg(y) \geq f(y)$ for any y. More precisely, T follows a geometric distribution with parameter $1/M$, and the expected number of trials until a variable is accepted is M. We will use this later. Finally, T is almost surely finite, meaning the algorithm cannot run indefinitely.

By the monotonicity of probability measures on an increasing sequence of events, we have:

$$\mathbb{P}(T = +\infty) = \mathbb{P}\left(\bigcap_{k=0}^{\infty} \{T = k\}\right) = \lim_k \mathbb{P}(T = k).$$

Since T follows a geometric distribution with parameter $\frac{1}{M}$, the probability of $T = k$ is given by:

$$\mathbb{P}(T = k) = \left(1 - \frac{1}{M}\right)^{k-1} \frac{1}{M}.$$

As $k \to +\infty$, this probability tends to zero:

$$\mathbb{P}(T = +\infty) = 0.$$

We could also state that $T < \infty$ a-e as the expectation of T : $\mathbb{E}(T) = M < \infty$, then $T \in L^1$ which implies the finitude of T.

Based on the values provided in the problem statement, we obtain the following graph for f (shown in purple) along with the instrumental density f_1 (shown in black):

This graph confirms that f_1 is indeed an appropriate choice for the instrumental density, as it envelops

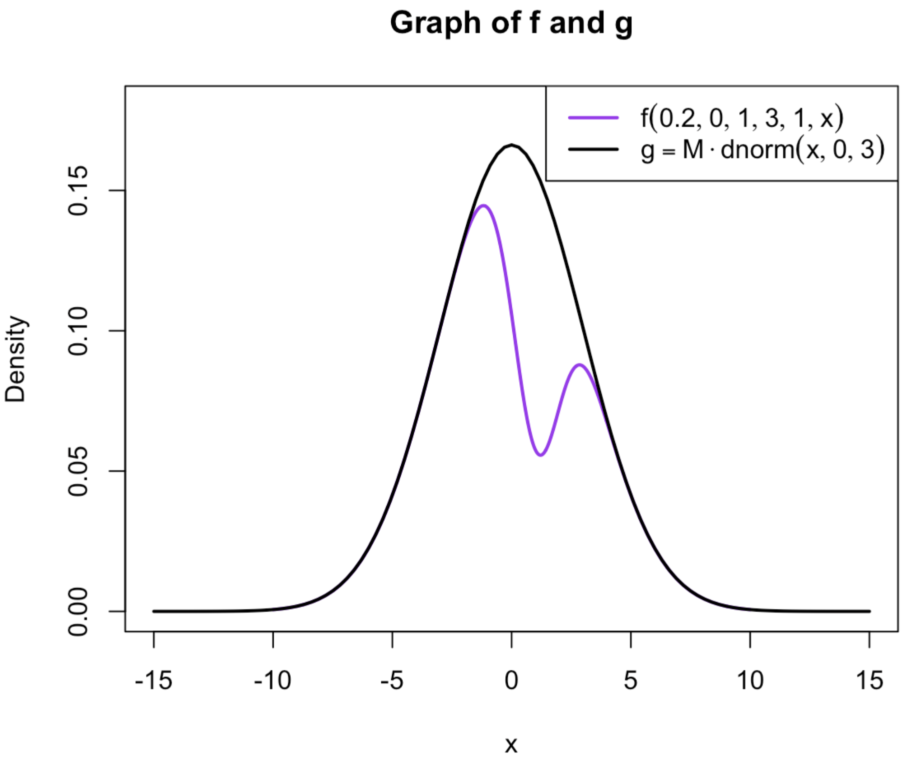

Figure 11: Graph of f and f_1 showing that f_1 is an appropriate instrumental density.

f effectively.

3.2 Question 7

Write a function `accept reject(n)` *that generates n samples from f using the accept-reject method. Generate* `n=10000` *samples, and graphically check that* `accept reject()` *is correct. Compute the empirical acceptance rate and check whether it agrees with its theoretical value.*

When running the code, considering the values given at the beginning of the exercise, we obtain this histogram b y generating $n = 10^4$ samples.

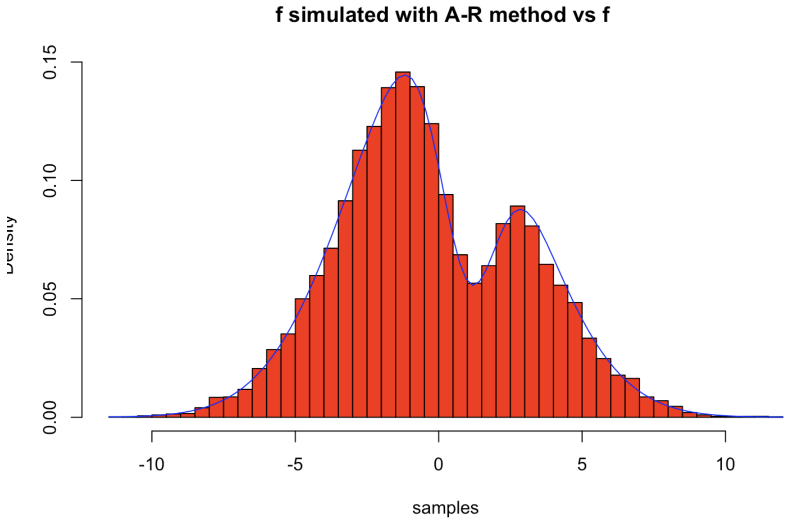

Figure 12: **Histogram of samples generated using the Accept-Reject method compared to the target density f. (10^4 samples).**

As n increases, the red plot tends to embrace more and more the curve of the real-non randomly generated f.
Then we want to check if the empirical acceptance rate agrees with the theoretical expected rate. As we computed before, $\mathbb{E}(T) = M$, then the expected rate should be 1.25. Indeed, we had $M = \frac{1}{1-a}$, with $a = 0.2$.
We prefered to plot many values of the empirical rate obtained. To this end, we had to add a "Loop"-counter in the accept-reject code, then extract it, then divide the amount of sample n by the the sum of the loop and n, which agrees with the theoretical background of the method, as the rate has to be less than 1, and also because $M > 1$ so that $\frac{1}{M} < 1$. Then we have the following histogram of empirical rates, showing that they are well centered and really close to the theoretical one : 0.8.

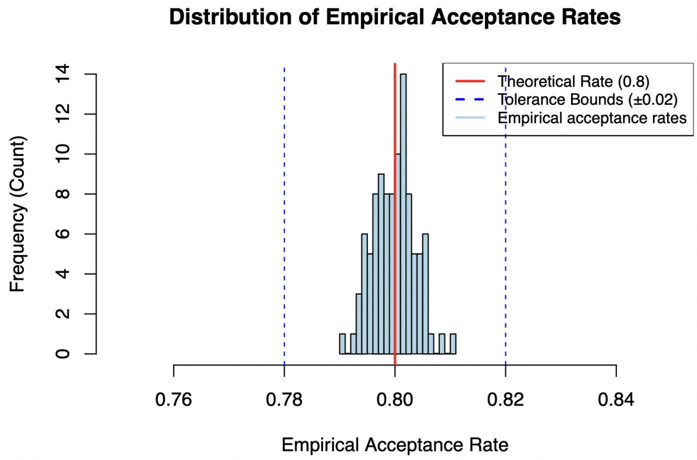

Figure 13: **Histogram of 100 empirical rates computed, well distributed in a neighborhood centered in 0.8 of arbitrary radius 0.02.**

3.3 Bonus for Question 7*

We wanted to add another way to find the empirical rate, as it has to be the expectation of the random variable T, we took example from an exercise from the TP sheet, the 2nd one, where we have to simulate N, defined also as an inf, which is a Poisson law, depending on the sum of iid-Exponential parameters. We use a matrix to store in each coefficient a couple of random uniform and gaussian normal, and we accept the Y normal when the condition of acceptation holds.
We then compute the mean of the vector T which takes the first integer such that the acceptance condition is checked, otherwise it takes the arbitrary value $10^4 + 1$, T may be considered as a truncated geometric variable.
We choose $n = 10^4$ rows because the probability that $T > 10^4$ is extremely low for a geometric distribution with parameter $p = 1/M$. With $M = 1.25$ (when $a = 0.2$), the probability is $\mathbb{P}(T > 10^4) = 0.2^{10^4}$, which is approximately zero.
We used the "which" operator as we are dealing with a huge squared matrix of size $10^4 \times 10^4$, it allows us to not use any "if" operator or "while" as this not fits well with R language.

```r
#We generate the matrices U, Y, and the values of T
result <- generate_matrix_and_T(n, M, a, mu_1, mu_2, s_1, s_2)
U_matrix <- result$U_matrix  # Matrix of U values
Y_matrix <- result$Y_matrix  # Matrix of Y values
T <- result$T  # Vector of T values for each row

mean(T)  # This should yield a value close to M
```

```
## [1] 1.2504
```

Figure 14: Empirical Value of M using the fact that it is the expectation of a Geometric random variable.

We found something really close to 1.25, thus we can be satisfied from this other way to find the empirical rate, as now we just have to compute 1/mean(T), also close to the expected theoretical value of 0.8.

3.4 Question 8

In this question, we decided to plot the values of our empirical acceptance rate, and compared it with the theoretical one. As we know, $1/M$ should be equal to $1 - a$ for $a \in]0; a*]$.
First, we compute the accept reject function significantly such that we have a "quasi-continuous" plot, easily comparable to the theoretical results. We make use of the "sapply' operator as we do not want to make any loop with "for" or "if".

We obtain :

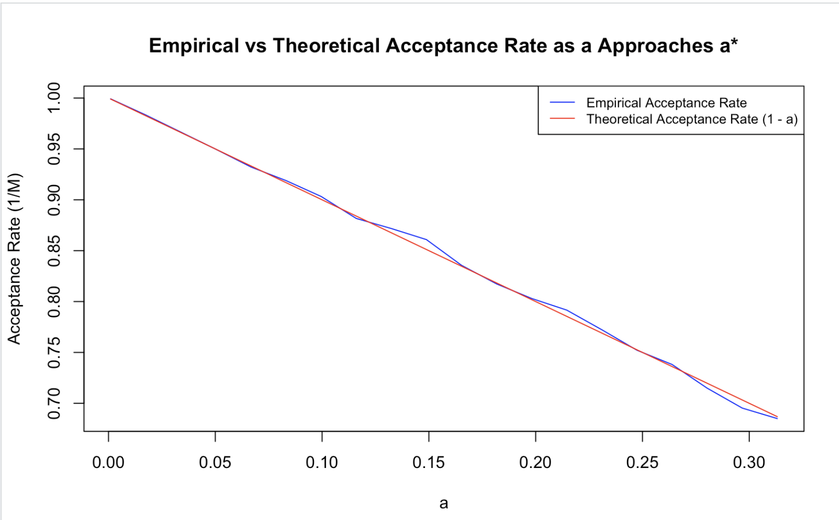

Figure 15: Empirical acceptance rate vs theoretical one, for $a \in]0; a*]$
.

As $a*$ is attained, we would like to plot the impact on the simulated f when we consider the final limit case $a = a*$:

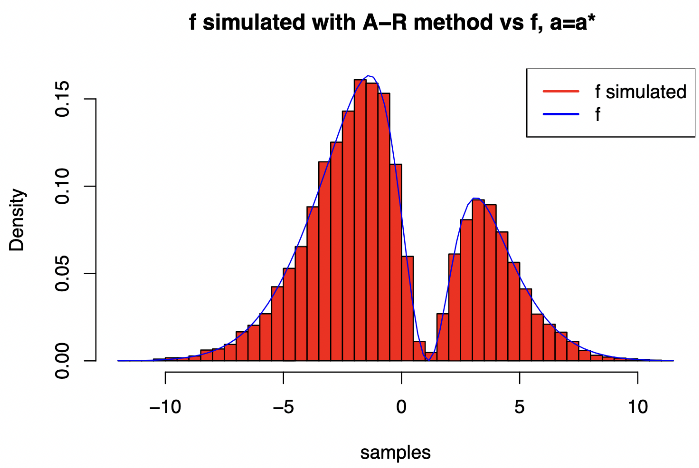

Figure 16: Simulation of f vs Non simulated, $a = a^*$
.

Clearly, the peak is maximal and we really see the two normal laws being well distinguished and making their own trend.
From the Figure 15, we clearly have that the acceptance rate is linearly decreasing as a tends to $a*$. This means that the amount of loops is higher and that we fall more frequently in the rejected zone. As the

rate is decreasing, the discrepancy factor M is converging towards $\frac{1}{1-a*}$ and is too high, which implies that the accept-reject method is the least efficient for this $a*$ and we may have to simulate with another method as it takes too much times and too much computations.