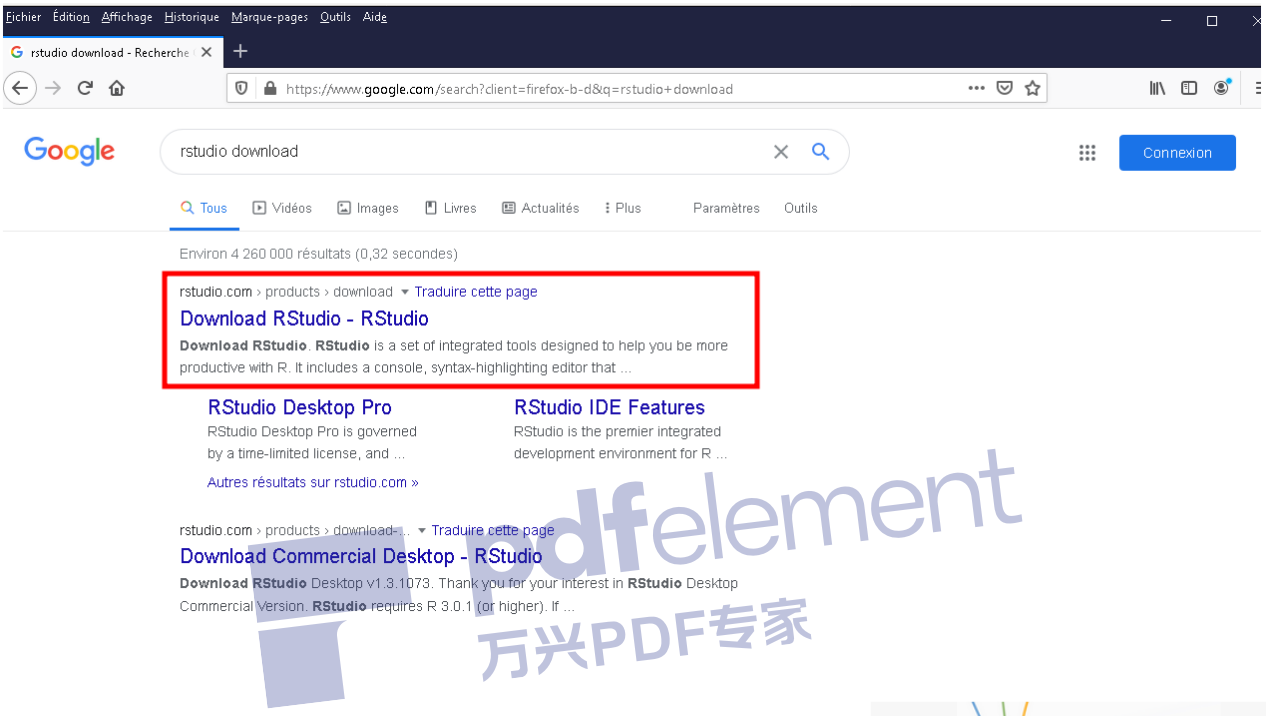
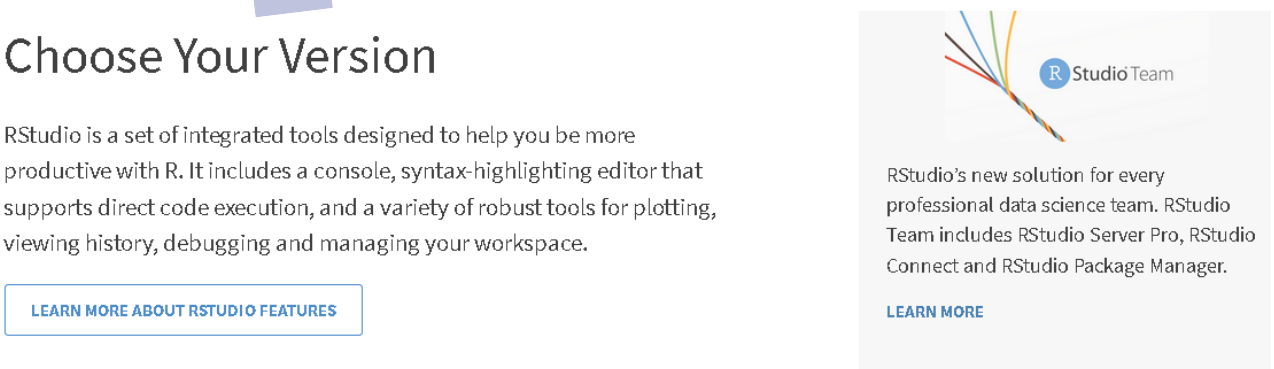
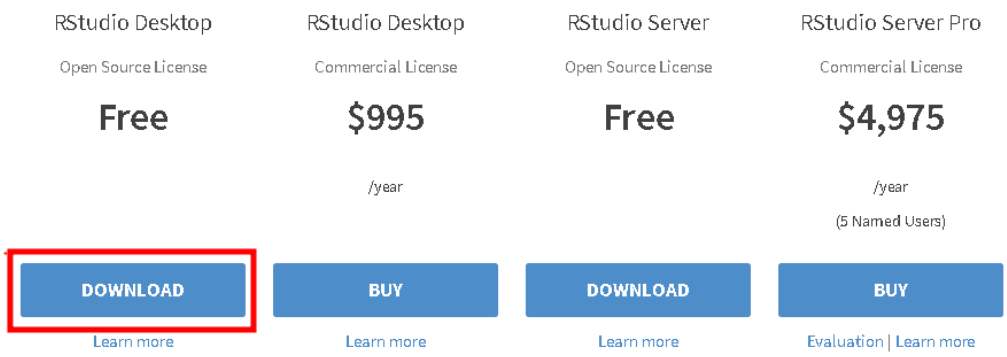
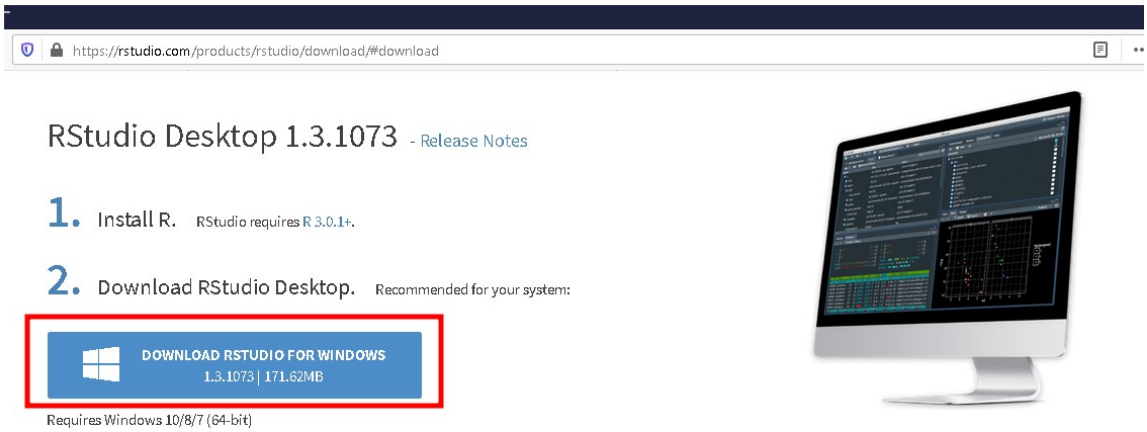
Practical Work I
Model comparison

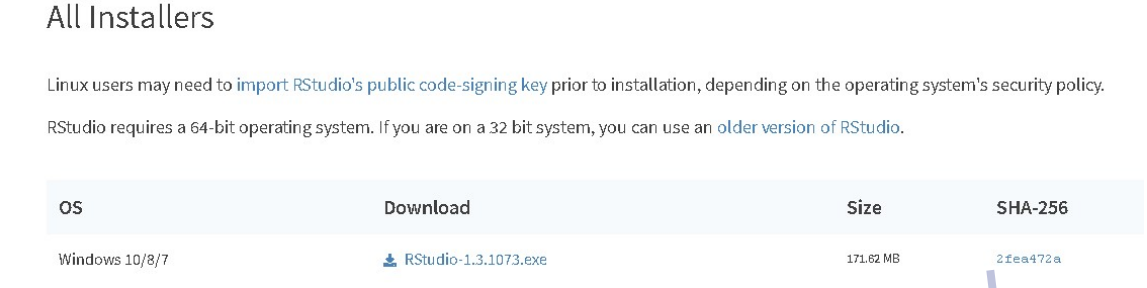
How to install and use Rstudio

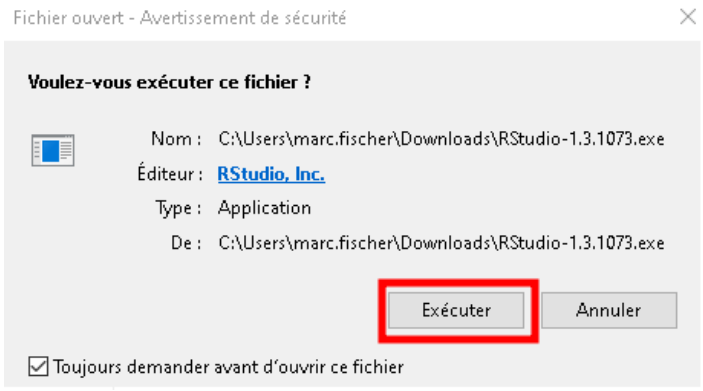












Install the program.

The ultimate goal of this practical work is to compare two reaction models of chemical kinetics with respect to one experimental profile through the use of **Bayesian reasoning**.

The two models are:

- M_1 : An isomerisation towards product P_1:

$$R \rightarrow P_1$$ 异构化

$$\frac{d[R]}{dt} = -k_1[R]$$

- M_2 : A recombination towards product P_2: 组合

$$R + R \rightarrow P_2$$

$$\frac{d[R]}{dt} = -2k_2[R]^2$$

Pay always a close attention to the units!

The units of k_1 are s^{-1}.
The units of k_2 are $mol^{-1}.cm^3.s^{-1}$.
The unit of the time is **µs** and those of the concentration are **mol/m^3**.

The initial concentration of the species is given by $[R_0]$ = 1200 mol/m^3.
The experimental profile is located in the file "**R_Exp.csv**".
We suppose that the experimental data are normally distributed with the following standard deviations $\sigma_i = \epsilon R_{\exp,i}$. ϵ is a constant. We'll first assume it is equal to **0.065**.

All along, we shall consider that $k_1 \in [2; 4.26E+3]\, s^{-1}$ and $k_2 \in [5E+04; 5E+09]\, mol.cm^{-3}.s^{-1}$

The program defines and produces three prior probability distributions. Each one of them is uniform with respect to one of the variables.
f_k(k) is uniform with respect to ***k***.
f_u(u) is uniform with respect to ***u = 1/k***.
f_w(w) is uniform with respect to ***w = log10(k)***.
Whenever you're asked to plot something, you must copy and paste the graphic into this Word document that shall be your report which must be returned to the instructors for an evaluation.

The practical works must be done and completed in groups of three or four. If you don't manage to finish it before the end of the practical work, please return them to me in the next five days.

Write the answers to the questions in that document in blue.

I. First model

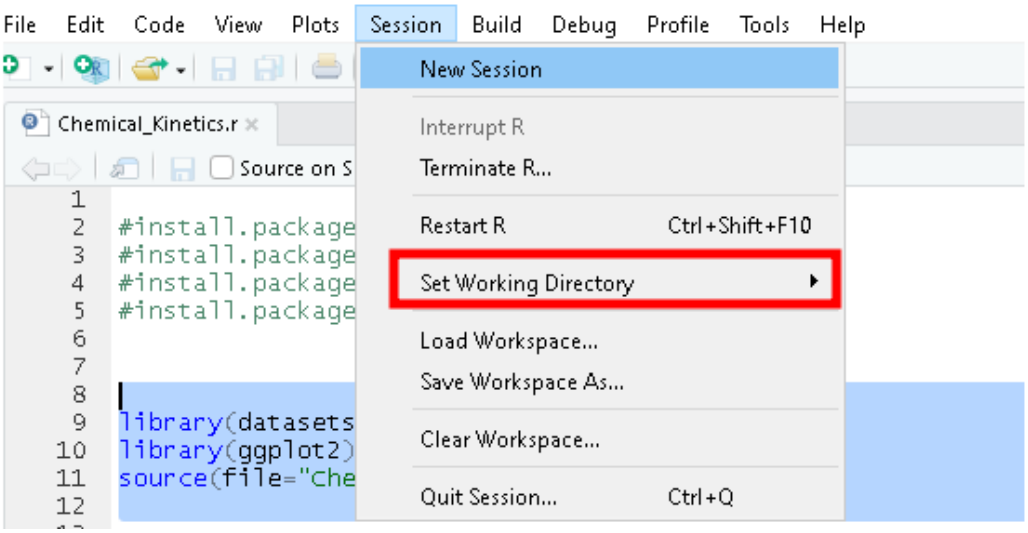
All functions necessary for carrying out the Bayesian analysis of model M1 are already present in "***Chemical_Kinetics_Functions.r***". Normally, you won't need to create any brand-new functions.

You'll call these functions in the file "Chemical_Kinetics.r" which you must now open with **Rstudio**.
You'll write the instructions to answer each question in both "Chemical_Kinetics.r" and the present Word document. This file too must be returned to the instructors.

Do not try to directly compile the function file "***Chemical_Kinetics_Functions_M1_windows.r***" "!

0) If it hasn't already been done, install the powerful plotting package "***ggplot***" by commenting out and compiling *install.packages("ggplot2")*.
Install also "***scale***" and "***gdata***".

Choose to work in the current directory:



Select "**To Source File Location**".

1) While these libraries are being installed, read carefully and thoroughly the file "**Chemical_Kinetics_Functions.r**".
Use the comments to try to understand what each function does.

This will be extremely useful for answering the questions without "reinventing the wheel".

2) Check and visualise the priors.
Verify that their integrals are always equal to 1.

Tips: you can enlarge the width of the plot if the axis ticks are too close to one another so that the numbers end up hiding each other.

3) Plot the profile for k = 1.357e+03 /s and compare it with the measurements.

4) Plot the likelihood and verify that the optimal point is approximately k = **1.357e+03** /s.
Focus on the region where the likelihood has the strongest values for producing the graph.

5) Compute the posterior probability distributions corresponding to the three priors.
Check that their integrals are equal to 1.
Represent them graphically. Choose an appropriate interval where the values of the likelihood (and therefore also of the posteriors) are the highest.
While representing the graphic, increase manually its width if the axis values aren't readable.

6) For each of the three priors, compute the prior probability that k belongs to the interval [1000;1500].

Are the three prior probabilities in relatively good agreement with one another?

7) Compute now the three corresponding posterior probabilities $p\left(k_1 \in \left[1000\,;1500\right] \middle| experiment\right)$
Tips: you could try to employ the same function as the one you used for the previous question.

read.table could also prove useful.

Can we say that the experimental data washed out the prior?

8) Compute now the three posterior probability distributions for $k_1 \in [2\,;4.26\text{E+}3]\,s^{-1}$ and $\epsilon = 30\,\%$.

Plot the three posteriors for $\epsilon = 6.5\,\%$ and the three posteriors for $\epsilon = 30\,\%$ beside one another while using the same scale.

9) Calculate now the three posterior probabilities $p\left(k_1 \in [1000\,;1500] \middle| experiment\right)$ for $\epsilon = 30\,\%$.
Compare them with those for $\epsilon = 6.5\,\%$.
Try to explain the differences by plotting the likelihood for $\epsilon = 30\,\%$ and that for $\epsilon = 6.5\,\%$ beside one another with the same scale.

II. Second model.

We're now considering the second model.

M_2 : A recombination towards product P_2:

$$R + R \rightarrow P_2$$

$$\frac{d[R]}{dt} = -2k_2[R]^2$$

We assume that $k_2 \in [5\text{E+}04\,;5\text{E+}09]\,mol\,.cm^{-3}.s^{-1}$ and $\sigma_i = \epsilon\, R_{\exp,i}$ with $\epsilon = 6.5\,\%$. We still have $[R_0]$ = 1200 mol/m^3.

1) Save "Chemical_Kinetics_Functions_M1.r" under "Chemical_Kinetics_Functions_M2.r" without erasing the first file.
Modify this new file so that everything corresponds to the second model M2.

2) Compute and plot the three priors.

3) Compute the likelihood.
Plot it in an appropriate, narrower interval.

4) Plot the optimal profile for M2.
Compare it to the optimal profile for M1.

5) Compute the three posterior probability distributions.

III. Comparison of model M1 and model M2

As we saw during the lecture, the ratio of the posterior probability of M_1 and M_2 is such that

$$\frac{p(M_2|E)}{p(M_1|E)} = \frac{\int L(E|M_{2,}k_2) f(k_2) dk_2}{\int L(E|M_{1,}k_1) f(k_1) dk_1} \frac{p(M_2)}{p(M_1)} = B \frac{p(M_2)}{p(M_1)}.$$

$B = \frac{\int L(E|M_{2,}k_2) f(k_2) dk_2}{\int L(E|M_{1,}k_1) f(k_1) dk_1}$ is the Bayes' factor which is supposed to represent the impact of the data on the probability ratio of the two models.

Let's suppose we applied the **principle of indifference** so that p(M1) = P(M2) = 0.5 and $\frac{p(M_2|E)}{p(M_1|E)} = B.$

1) Compute B_k for the **first** uniform prior on *k* but not for the two other priors.
Replot the two optimal profiles of *R(t)* beside each other.

What does the value you get imply for model M1 and model M2?

Does that make sense to you?

2) Compute now B_k, B_u and B_w at the same time.
Are the results internally consistent?
What do the values of B_u and B_w mean for the models M1 and M2?

3) Propose an explanation for the paradoxically small value of the Bayes' factor **B_k** corresponding to the first prior. You should base your reasoning on the formula defining **B**.