The present document aims to estimate the « mortality rate » of CoVid 19.

A preliminary warning is required about the expression « mortality rate » I used, and will use all along this note.
In https://ourworldindata.org/coronavirus#what-do-we-know-about-the-risk-of-dying-from-covid-19 three different indicators of mortality are discussed :

- the Case Fatality Rate (CFR)
  defined as the number of deaths from desease divide by the number of diagnosed cases of desease,

- the Crude Mortality Rate (CMR)
  defined as the probability that any individual in the population will die from the disease; not just those who are infected, or are confirmed as being infected,

- the Infection Fatality Rate (IFR)
  defined as the the number of deaths from a disease divided by the *total number of cases* (not the number of the diagnosed ones).

The « mortality rate » I refer to is a mix between CFR and CMR : it assesses
the raw probability that someone diagnosed positive will die from the disease.
where **raw** means here « without any distinction of age, gender, health status or anything else ».

This « mortality rate » (I will drop the quotes in the sequel) is obviously larger than CMR (which is itself larger than IFR). One can also prove it is larger than CFR as long ast he outbreak is still in progress (the reason is given below but I give here a short hint why : the number of persons that die some day has to be compared to the number of persons that where diagnosed positive days before).

The indicator I present is probably more interesting for the common man: "How many chances do I have of dying if I am detected infected? "than it is for epidemiologists.

Whatever the indicator we adopt, it is a general rule that assessing it as the outbreak is still on progress is a very difficult problem, see

for instance *https://academic.oup.com/aje/article/162/5/479/82647* .

At the contrary its solution is obvious as soon as the outbreak is over (at least for a factuel indicator such as CFR) . Nevertheless some « subjective » indicators (CMR and IFR) need to know some numbers we can only guess, for instance what is the proportion of people that passed under the radars ? And even, one century after the spanish flue, specialist still do not agree about its true mortality.

At the world level, CFR was estimated to 3.6 % on march 26.

Another method of calculation is based on the fact that a person diagnosed positive has only two possible outcomes: either it recovers either it dies. If D(t) is the cumulative number of deaths at date t and R(t) is the cumulative number of recoveries at the same date, then the mortality rate is estimated by the formula D(t) / (D(t)+R(t)). It is easy to show that this value, if the mortality rate is stationary (i.e. the typology of people who die does not change over time and no mutation of the virus occurs), overestimates the actual mortality rate. To date, at the global level, this estimate provides a mortality rate of about 16% (including all genders, ages and people with aggravating factors of comorbidity).

The reality is probably, or probably not, between these two numbers.

If you are interested in this subject here is a site where all this is fairly well explained:
https://www.worldometers.info/coronavirus/#repro or, directly for the mortality rate (called CFR for Case Fatality Rate):
https://www.worldometers.info/coronavirus/coronavirus-death-rate/

Recently a controversial article in the Lancet (the leading medical journal) reported some rather pessimistic estimates
https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(18)30537-0/fulltext
The calculation is based on the following simple observation: people who die on date t are those who were infected on date t-T.
If c(t-T) is the number of new cases detected on date t-T and d(t) is the number of new deaths on date t, then the mortality rate is simply equal to d(t)/c(t-T).
It is clear that the greater T is, the greater this rate is during a period of exponential growth of the epidemic. A simple summation over time shows that if C(t-T) is the cumulative number of confirmed cases at date t-T and D(t) is the cumulative number of deaths at date t, then the mortality rate is equal to D(t)/C(t-T).

All the data I use below are available here https://ourworldindata.org/coronavirus (csv files for D(t) and C(t) for the different contaminated countries).

The Lancet's article is questionable because :

- T is not known for certain
- More important is that the survival time (T) between a positive diagnosis and the death is not a deterministic quantity one can reduce to a single delay T, but a random variable which depends on :
    - the age,
    - the gender,
    - the health of the individual,
    - the health care system,
    - the infection detection policy (currently a controversial issue)
    - ... and on the hazard

The idea developed here is to use a Bayesian Inference (BI) procedure https://en.wikipedia.org/wiki/Bayesian_inference to jointly estimate the mortality rate and the survival time (more precisely to assess the joint posterior distribution of both of them).

I won't go into detail about the process which underlies to this inference, but I will just give you an outline of it.
In BI one consider that the degree of belief about any unknown quantitiy can be represented by an ad hoc random variable whose distribution is terme « prior distribution » random variables.
The case of the mortality rate is rather simple: in a complete the lack of solid information we can consider a uniform random variable on]0, 1] correctly represents this prior belief. If stronger informations are avaliable, one could fir instance use a uniform distribution over [0.04, 0.08], or a gaussian distribution, or any other .
The case of the survival time T is far more complex.
From https://www.worldometers.info/coronavirus/#repro its read that (this depends of the age group) it range from 6 to 41 days with a median of 14 days.. How can we model our degree of belief about T given these imprecise data ?
Among the many possible ways there is this one :

- Let X a random variable of Beta type with distribution Beta(2, 6)
- Then T = 6 + (41-6) * X is a random variable with support [6, 41] and median 14.

One problem here is that these numbers 46, 41 and 14 depend on many parameters, amid them are :

- the group age
- the country
- the health care system
- ...

More of this one can find many other distributions with support [6, 41] and median 14 (for instance a triangular distribution, or a correctly truncated gaussian distribution to mention a few).
A classical way to proceed in BI is then to say that the « shape » of this distribution is itself an unknown quantity one can model by an ad hoc random variable .
This approach is termed « Hiererchical Bayesian Modeling » (HBM) https://astrostatistics.psu.edu/RLectures/hierarchical.pdf .
In the present case it consists in choosing a family of distributions whose parameters are capable to handle very different shapes od probability density functions (pdf). A common candidate is the Beta familly.
Let again X a random variable with distribution Beta(α, β) : if $\alpha = \beta = 1$ one obtains a uniform distribution, if $\alpha = \beta >> 1$ the pdf looks like a gaussian's one centered at ½, for $\alpha < \beta$ ($\alpha > \beta$) one obtains a negatively (positivey) screwed distribution, and even an exponentil like distribution if $\alpha < 1$ and $\beta > 1$. Thus a Beta distribution appears to be a very versatile one.
Doing this introduces de facto two new quantities α and β we know nothing about. The two quantities are dubbed « hyperparameters » https://en.wikipedia.org/wiki/Hyperparameter.

Here again the degree of belief of each of them is modeled by an ad hoc random variable. The simplest solution is to use a uniform random variable over some range [$p > 0$, $q > p$] (of course we could add another layer in the HBM approach but it would probably add here an unnecessary complexity).

To close this probabilistic we need to define th a priori relations among these 5 random variables (from now on I will use a bold font to denote a random variable and a normal one to represent any of its realization) $\boldsymbol{\tau}$, $\mathbf{a}$, $\mathbf{h}$, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$. (from now on $\boldsymbol{\tau}$ will denote the mortality rate).
Without any further information the most reasonable choice is to consider that thes 5 random variables are mutually independent.

The final model then is (the symbol « ~ » means « leads a distribution », the values of the parameters of the distributions are illustrative and will be adjusted for the intended BI [see the mw file]):

- τ ~Uniform(0, 1)
- $\mathbf{a}$ ~DiscreteUniform(3, 8)
- $\mathbf{h}$ ~DiscreteUniform(3, 30)
- α ~Uniform(1, 6)
- β ~Uniform(1, 6)
- $\mathbf{U} = (\tau, \mathbf{a}, \mathbf{h}, \alpha, \beta)$, $\mathbf{U}_m \perp\!\!\!\perp \mathbf{U}_n$ for each couple $(m, n>m) \in [1..5]^2$ (the symbol « $\perp\!\!\!\perp$ » is an usual statistical symbol to denote independancy)
- $\mathbf{T} \sim \mathbf{a} + \mathbf{h} \times \text{Beta}(\alpha, \beta)$

We therefore have to solve an inference problem on 5 unknown parameters.
The initial problem of infering the mortality rate (more precisely the posterior distribution of τ) has been complicated by adding 4 extra parameters $\mathbf{a}$, $\mathbf{h}$, α and β. From the point of view of the sole inference about τ these extra parameters are dubbed « nuisance parameters » https://en.wikipedia.org/wiki/Nuisance_parameter. T
his term of « nuisance » expresses the fact that the ignorance of the real values these parameters have fuzzifies the proper estimation of τ. What we must expect is that the introduction of these nuisance parameters will result in a less well located and more spread posterior distribution of τ. This is the price to pay to obtain a more reliable inference when only loose prior informations are avaliable.

Basicaly BI consists in assessing the quality of a realization U of $\mathbf{U}$ by comparing observations and predictions. More precisely this means computing three quantities :

1. The prior density of probability $\pi(\mathbf{U}=U)$ simply noted $\pi(U)$
2. The likelihood of the observations under the hypothesis that $\mathbf{U}=U$ is the true choice to make
   This term will be denoted $L(O \mid U)$ https://en.wikipedia.org/wiki/Likelihood_function

3. The posterior density of probability of U $Po(U) = L(O \mid U) \times \pi(U)$

As a rule $L(O \mid U)$ depends on the statistics of an observation error, for instance a measurement or an experimental error. Typicaly this error reflect the fact the doing again the same experiment results in a new observation which differs from the previous one by some « random effect ».
Obviously this kind of measurement does not exist here, for the recorded number of deaths is, at any time and for each country, a perfectly known number.
One classical way to manage the lack of the likelihood (or the fact it can be amazingly expensive to compute it) is to use an approach termed ABC (Approximate Bayesian Computation https://en.wikipedia.org/wiki/Approximate_Bayesian_computation).

The approach I used here is a mix of ABC and MCMC approaches termed ABC-MCMC or « likelihood-free MCMC », here is one of the best talk about BAC methods http://www.maths.lu.se/fileadmin/maths/forskning_research/InferPartObsProcess/abc_slides.pdf .
ABC-MCMC require defining a « distance » between observation and simulation that I will name below « pseudo-likelihood ».
Before explaining what this pseudo-likelihood is, some details about the inference algorithm must to be explained.

- Let c(t) the number of confirmed cases recorded at time t (in fact t takes only discrete values because reports are given day by day ; then t is not really the « time » but more precisely the « number of days from some reference date »)

- Let d(t) the number of deeath recorded day t.

- Let U some realization of $\mathbf{U}$:

    - From on $\tau \equiv U_1$ one can assess the number of people confirmed positive at date t that are goind to die.

      This number d'(t) is a random realisation of a binomialrandom variable with parameters c(t) and τ : $d'(t) \sim \text{Binomial}(c(t), \tau)$.
      d'(t) is computed from all the recorded times t=0 from t=today.

- Let T some day between 0 and today.
  For each individual from the d'(T) ones a survival delay is chosen according by sampling the random variable

$T(U_2, ..., U_4) \equiv T(a, h, \alpha, \beta) = a + h \times Beta(\alpha, \beta)$.

We then know the date of death of each individual diagnosed positive at day t and who are going to die (just add its survival delay to the current day t).

- Doing this for each time from 0 to doday produces a « trajectory » of the number of deaths at any time.
  But.. one must observe this trajectory is not the result of a deterministic computation (as it often happens in the field of Bayesian Calibration of computer codes ; among the vast litterature on this topic see for instance http://www.int.washington.edu/talks/WorkShops/int_16_2a/People/Bingham_D/Bingham.pdf).
  Indeed two elements make this trajectory « random » : the fact the numbers d'(t) are random realization of a random variable, the fact that the survival delays are too.
  Mathematicaly this « trajectory » is thus a realization of a time-discrete stochastic process S https://en.wikipedia.org/wiki/Stochastic_process parameterized by U.
  Denoting as usual the hazard by ω, this trajectory can be denoted $S(t, \omega \mid U)$ where « | U » is a short for « given that **U**=U ».

- We build N independant realizations of S and assess its mean $\mu_S(t \mid U)$ by averaging out these « trajectories ».

  A standard deviation $\sigma_S(t \mid U)$ is computed simultaneously.

  Doing this I simply replace the outputs of the probabilistic model by summary statistics, a rather common way to act when these outputs are random realizations.

- One then assume S is a gaussian process https://en.wikipedia.org/wiki/Gaussian_process .
  Remark this process can't be a stationnary process for its expectation is a function of time. More of this its standard deviation is also a function of time (if only because the variance of Binomial(c(t), τ) depends on c(t) and thus on t).

- Let $\delta(t \mid U) = \exp\left(- \frac{1}{2} \times \left((D(t) - \mu_S(t \mid U)) / \sigma_S(t \mid U) \right)^2 \right) / \sigma_S(t \mid U)$

  A natural « pseudo-likelihood » *PL* is defined by the product of all $\delta(t \mid U)$ for t from 0 to today.
  Some precautions have to be done. For instance d'(t) can be null if the τd(t) is less than 1 (remember d'(t) ~ Binomial(c(t), τ)).
  In this case its likely that $\sigma_S(t \mid U)$ will be null for this same time. To avoid this situation I replaced $\sigma_S(t \mid U)$ by $\max(1e\text{-}6, \sigma_S(t \mid U))$.

- Note that the procedure PseudoLikelihood in the mw file contains other expressions for this« pseudo-likelihood » (last calling parameter of its interface) including :
    - The « δ(t | U) » criterion given above (option SelectedLikelihood:= 1)
    - **A weighted residual sum of squares** (option **SelectedLikelihood:= 2**)
    - A weighted variant of the « δ(t | U) » criterion (option SelectedLikelihood:= 3)
    - A weighted chi-square-like statistic (option SelectedLikelihood:= 4)

  Simulations tend to show a simplest weighted residual sum of squares is the best choice.

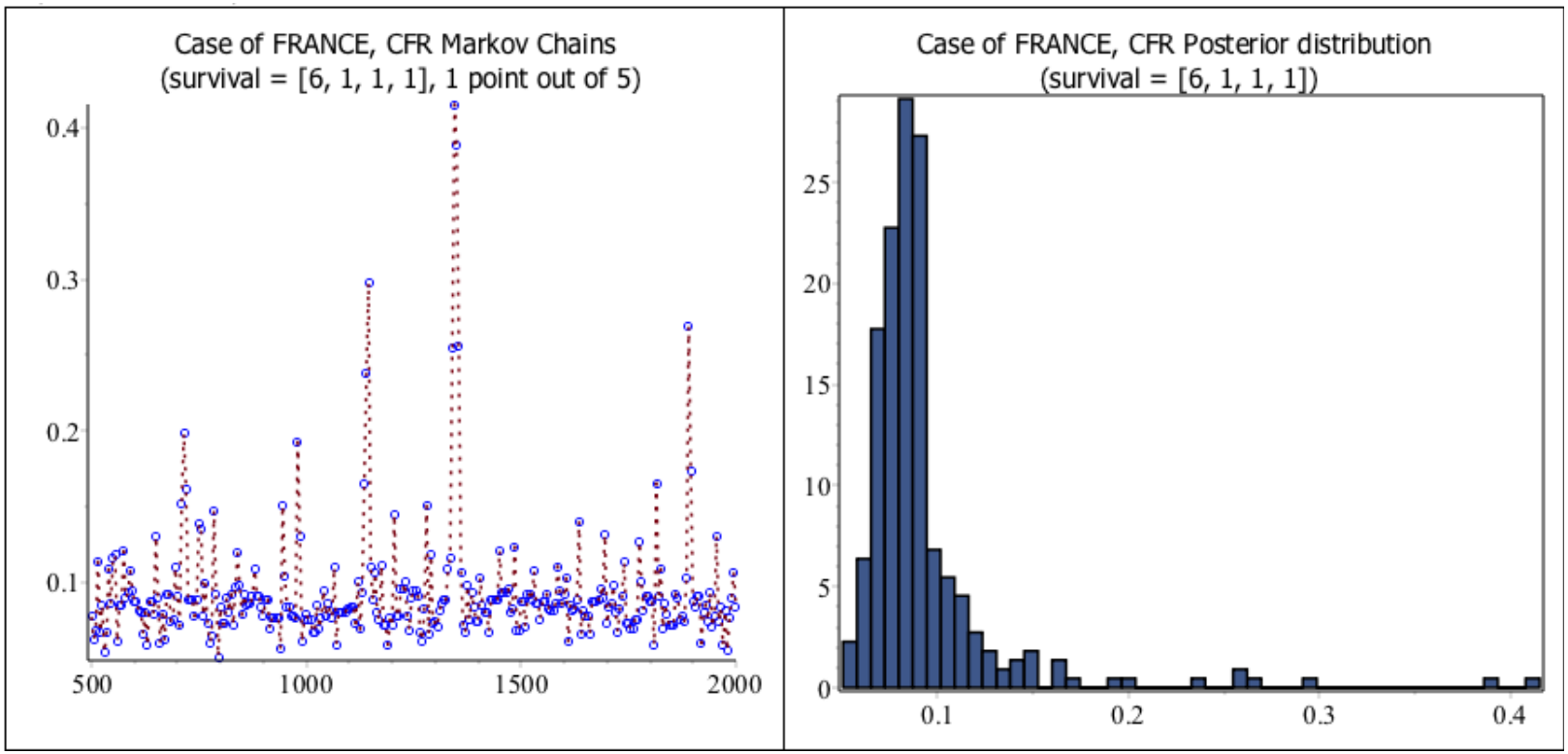
Nevertheless this is not the approach I used here. I used instead a specific procedure based on the definition of an ad hoc « pseudo-likelihood ».
Once a « pseudo-likelihood » has been chosen the inference problem is solved using a MCMC (Markov Chain Monte Carlo) method https://en.wikipedia.org/wiki/Markov_chain_Monte_Carlo .
I use here an ad hoc adaptation of the Metropolis-Hastings (MH) algorithm (see for example https://en.wikipedia.org/wiki/Metropolis–Hastings_algorithm) :

- at each step of the chain only one of the 5 parameters is allowed to change, its selection is done randomly,

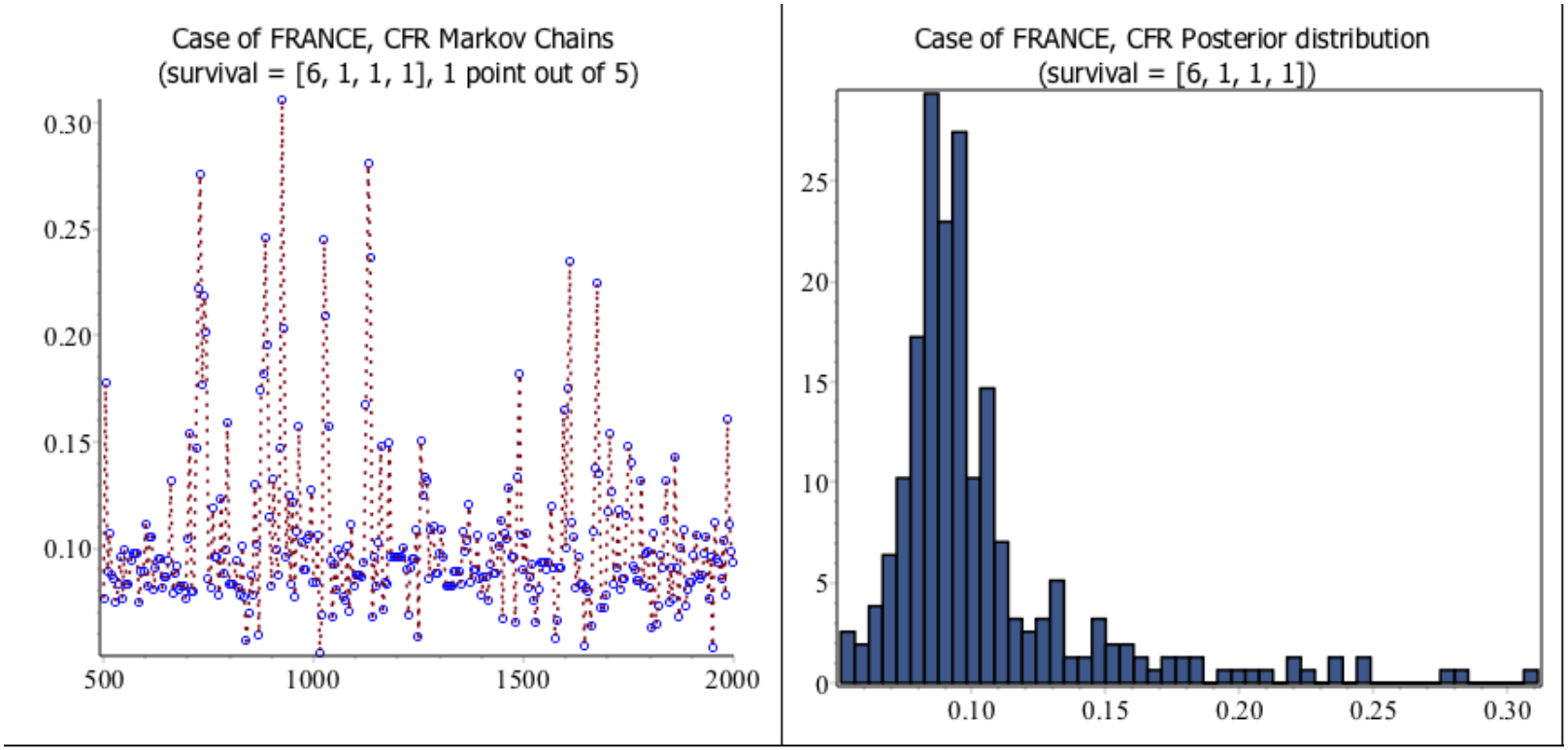
- only guesses within a the closed domain specified by the priors are generated (remember the priors ar uniform and thus bounded).

Here are some results obtained in the case of France (data come from https://ourworldindata.org/coronavirus ; the data used here range from decembre 31, 2020 to march 21, 2020 [more recent data are provided in the post]). They are unrealistic an presented here in order to help understanding what will become later ?

In these first serie of results **a**, **h**, α and β are assumed to be deterministic quanties ; the notation survival=[n1, n2, n3, n3] means **a**=n1, **h**=n2, α=n3, β=n4. Nevertheless the survival delay **T** is still a random variable.
The choice survival=[6, 1, 1, 1, 1] means the survival time is which can take values 6 or 7 equally likely (giging a median value to 6.5). This choice is an extremely informative... and unrealistic prior https://en.wikipedia.org/wiki/Prior_probability . The reason can be easily understood if we observe that this choice mimics that hypothetic scenario : as soon as a person is diagnosed positive at day t, a killer tosses a fair coin and kille this person on the date t+6 it he has got a tail or on date t+7 if he has gor a head.
In this case, the posterior distribution of the mortality rate, whose the prior distribution was taken uniformly on [0, 1] (note this prior is most non informative we can find) quickly becomes very prickly around 0.08 (= 8%).
These results have been obtained after launching two independent Markov chains (lengths equal to 2000, burn-in phase = 1...500; "1 point out of 5" means that only one point out of 5 is represented.
About the burn-in look for instance here http://users.stat.umn.edu/~geyer/mcmc/burn.html (this is an commonly used but controversial tool).

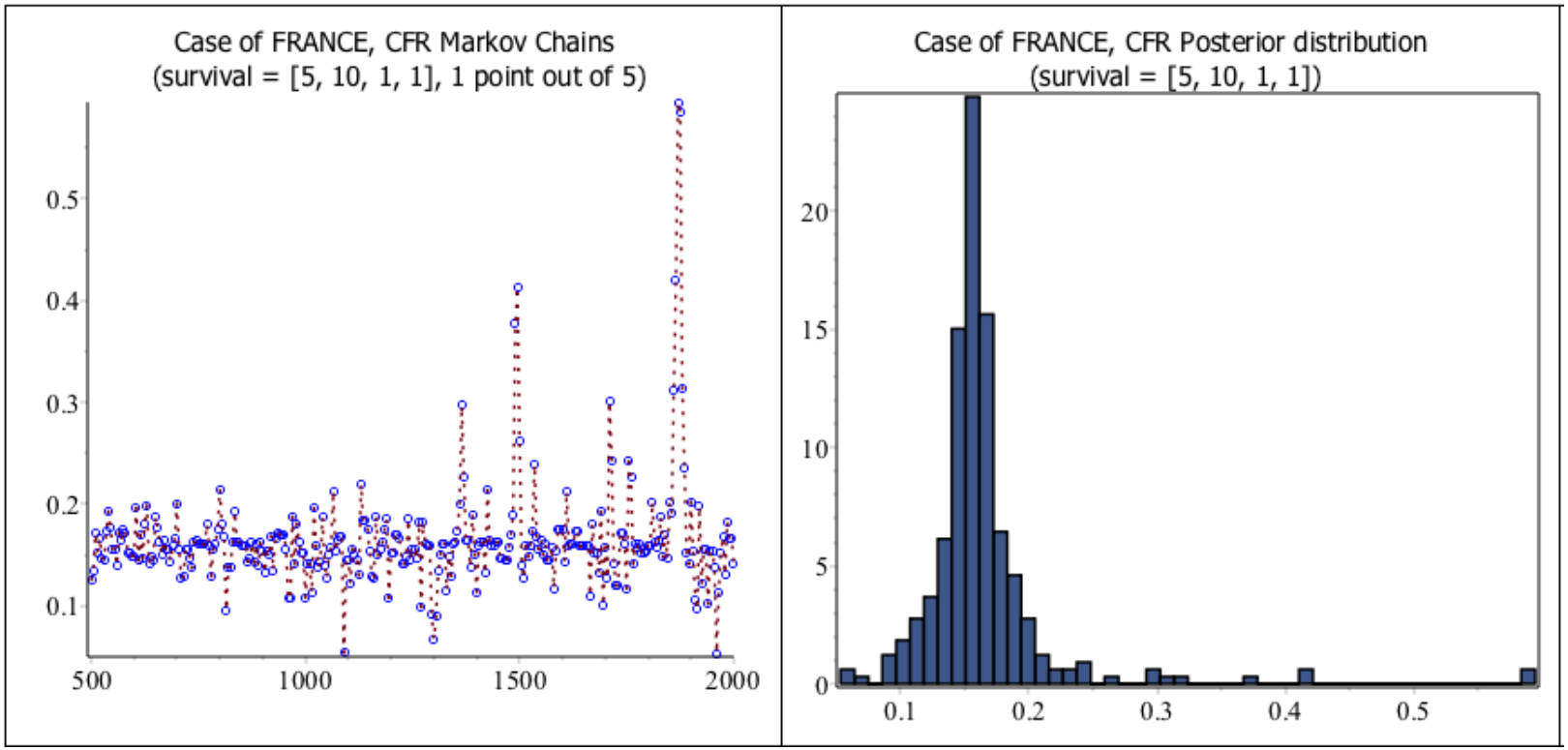


To understand the major role of survival time plays n the "online" (i.e. while the epidemic is still expanding) inference about τ, here is the result obtained with a slightly less informative prior for this survival time.
The choice now is survival=[5, 4, 1, 1] and results in a median survival delay of 7 days (almost the same as the one obtained with the previous survival choice). The death rate is now around 9.5%, and thus significantly higher than the estimate obtained above.
The 40% of people who died (in the simulation), before these 7 days contribute to lower the mortality rate estimate below the 8% above, while the 40% who died after these 7 days contributed to raise this rate. As the epidemic in France is growing exponentially (if you plot C(t) against t you will see that the curve is convex) more people contribute to raise the death rate above the previous 8% than to lower it.
Note also that, as the prior information on survival time is now more vague, the posterior distribution of the mortality rate is more extended towards higher values: the uncertainty on this survival time "harms" the estimation of this rate.



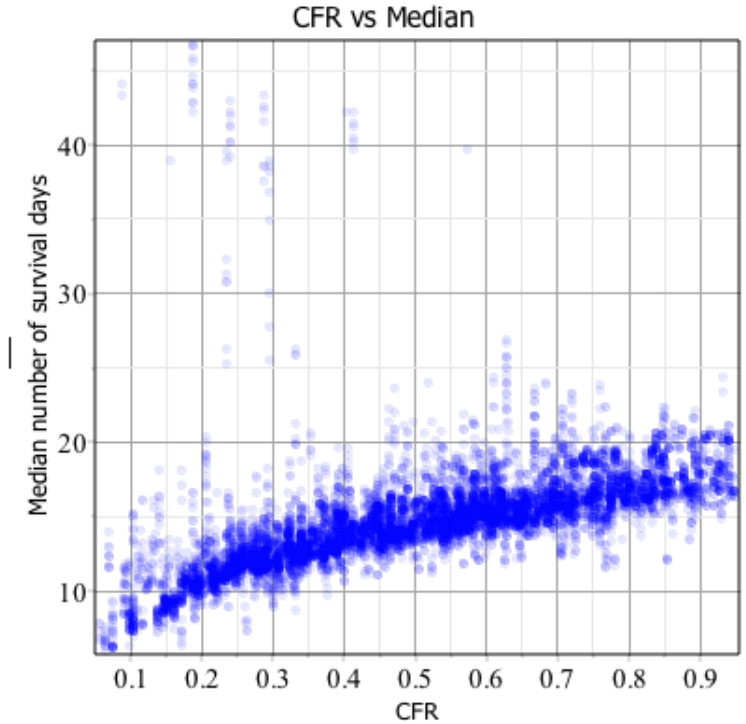
Finally let's take the even less informative prior survival=[5, 4, 1, 1] and. Its median has value 10 and corresponds roughly to the value used in the Lancet https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(18)30537-0/fulltext where this median was taken as a deterministeic resumé of the survival time **T** (here the killer arrives at your home exactly 10 days after you have been detected positive). The bayesian inference now gives a mortality rate concentrated around a value of 16% consistent with the ones given in the Lancet's.

Note that the approach of the author of this Lancet's article, even though widely criticized after its release, is now mentioned in https://www.worldometers.info/coronavirus/coronavirus-death-rate/ since a few days as a possibility to estimate « inline » the mortality rate.



This plot is aimed to show how the estimation of the mortality rate depends on the median of the survival time (of course this median cannot represents the dependency of τ to $\mathbf{T}$ in all its details!!!). To produce this plot realized different simulations of the type above by taking

- τ ~Uniform(0, 1)
- $\mathbf{a}$ ~DiscreteUniform(3, 10)
- $\mathbf{h}$ ~DiscreteUniform(3, 50)
- α ~Uniform(1, 10)
- β ~Uniform(1, 10)



A few caveats about the scope of this work:

- First of all it is not the mortality rate of the CoVid19 that is estimated here but "**the mortality rate of citizens of a given**

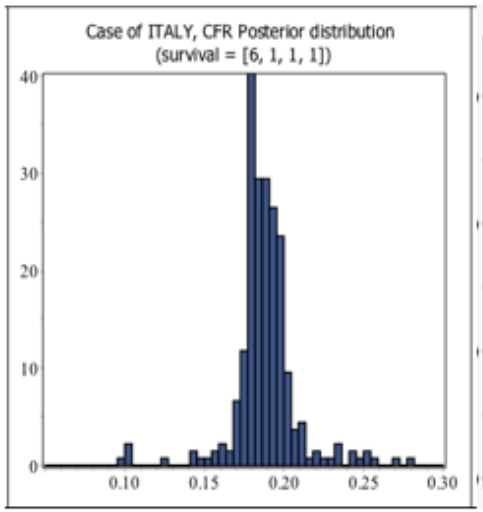
country that habe been diagnosed positive regarding the testing policy used". The bold terms are of great importance because, if the whole population of this country were tested every day, we would know exactly the number of confirmed cases, whether they were healthy, symptomatic, mildly symptomatic or severe carriers and thus, the number of deaths, which would probably be of the same order as that observed with the current testing policy, would mathematically lead to a lower mortality rate.

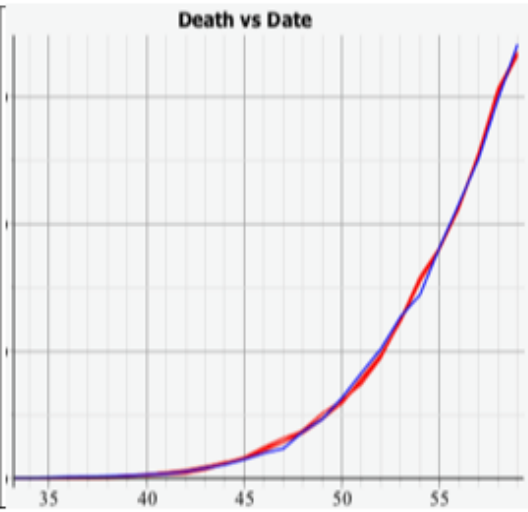
- Indeed, this is the main argument used in the media to justify a mortality rate of the order of 1 or 2%. But this is speculation about the number of people who pass under the radar. Even if the choices I have made are open to criticism (the a priori, the definition of a "pseudo-likelihood" that is calculable, ...), the values I obtain are objective and conditional to these choices. Beyond that, it is always possible to define other distributions a priori, other algorithms (ABC for example, see above).

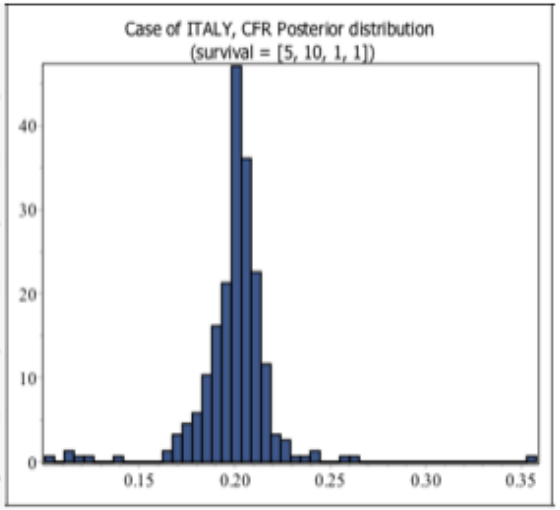
- The fact that all countries do not adopt the same testing policy leads to non-homogeneous samples from one country to another. In particular, it is not realistic, even if I have done so, to estimate with the previous procedure a global mortality rate.

A second example: the case of Italy

Case of Italy, second graph: in blue the observed curve the cumulative number of deaths, in red the predictions corresponding to the MAP (Maximum A Posteriori https://en.wikipedia.org/wiki/Maximum_a_posteriori_estimation , i.e. the best U* obtained with 3 independent Markov chains). The situation corresponding to this graph is survival=[6, 1, 1, 1, 1] .
Observe the estimated value of the mortality rate (around 20% for a median survival time equal to 7).







Analysis: The 2 histograms shown here for Italy differ radically from those obtained for France.

There are several reasons for this:

- The Italian population is often referred to as older than France's and elderly people are known to have a higher lethality rate than younger ones.

- The testing strategy.

- The hospital structure: see Libération https://en.wikipedia.org/wiki/Libération dated 23 March: "The Lombardy health system, victim of its excellence" in which Professor Giorgio Palù explains that in Lombardy (main source of infection in Italy, nearly 80% of deaths), the policy is the systematic hospitalisation of patients (coronavirussed or suffering from other pathologies) whose collateral effect has led to a very large number of nosocomial infections by SARS-Cov2 : "The Lombard's mortality rate of around 16% does not mean much".

Full Bayesian Inference : the example of France

When I first treated the case of France I used an extremely simplified version of the full problem by setting the four parameters **a**, **b**, α and β to fixed values. More of this I focussed the inference on the sole mortality.

I present now a few results obtained from a full bayesian inference on the same data used previously for the hrench example.

The prior I choosed is given below :

- τ ~Uniform(0, 1)
- **a** ~DiscreteUniform(3, 8)
- **h** ~DiscreteUniform(3, 10)
- α ~Uniform(1, 3)
- β ~Uniform(1, 3)

A few words about the choices of last four parameters seem useful :

- The choices for **a** and **b** may seem too informative. In fact I had used larger choices in a first attempt but a few Markov Chains randomly initialized showed that the ranges given here where reasonably non informative. From a technical point of view it's interesting to use reasonably large priors in order to avoid an excessive exploration of the prior domain and thus spare computational time.

- This same onservation also applies to the hyperparameters α and β.
- The survival time **T** then is a random variable with support [6, 18] and, I agree, the upper value 10 can be considered as being too small. But here again this is the consequence of the analysis of a preliminary work done with the larger choice
    - τ ~Uniform(0, 1)
    - **a** ~DiscreteUniform(3, 10)
    - **h** ~DiscreteUniform(3, 50)

- α ~Uniform(1, 10)
- β ~Uniform(1, 10)

- Last but not least : I used essentialy uniform distributions which, without going too deep inside the bayesian frameworks, means I adopted a particular stance one can qualify as « subjective bayesianism » see for instance
  https://en.wikipedia.org/wiki/Bayesian_probability#Objective_and_subjective_Bayesian_probabilities
  https://plato.stanford.edu/entries/epistemology-bayesian/
  https://hal.inria.fr/inria-00071367/document

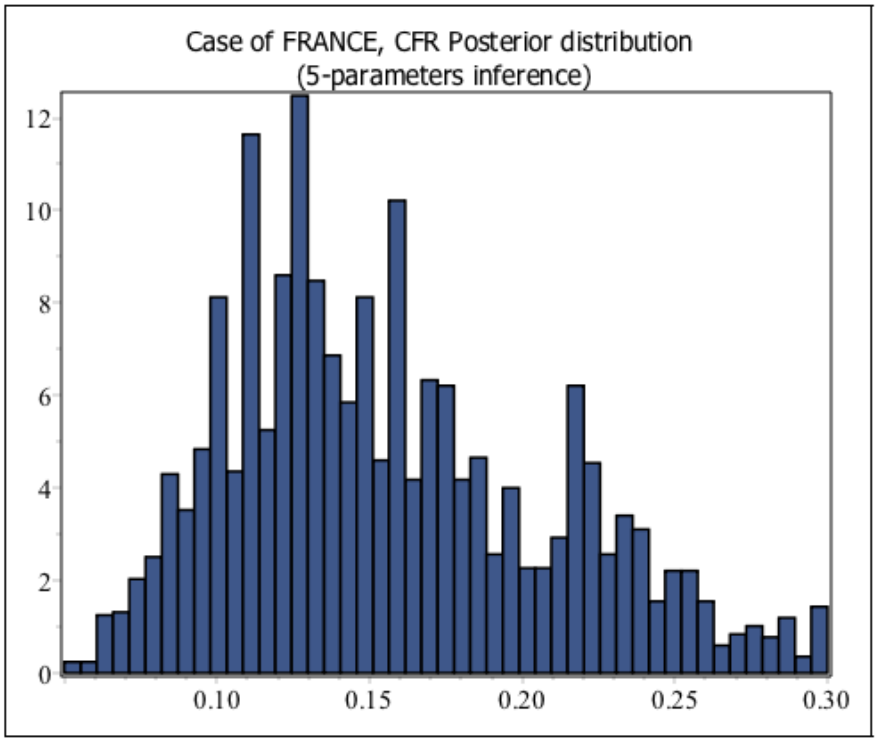
  This subjective stance is questionable from a philophical point of view (not from a technical one) and other subjective bayesians could have adopt other priors. At the opposite objective bayesians chose prior on the sole ground of pure mathematical arguments. So why did I used these priors ? Because I did this choice which seemed quite reasonable to me. But anyone else is free to use its own subjective priors given its own conviction… or even objective priors if it is objectivist.
  So no rule here.

Let Post(**U**) the posterior distribution of the 5-uplet **U**. The results below come from twelve independent Markov chains randomly initialized, each of length 2000 and with a burn-in range 1..500.
Some would say these lengths are too short and I agrre on this. This severe limitation is only due to the computational time used to draw longer chains (more precisely I tried to do my best in preventing the increase of the memory as the chain develops [which is the main factor which slows the computation] but I'm far to have enough skill to produce efficient Maple coding).

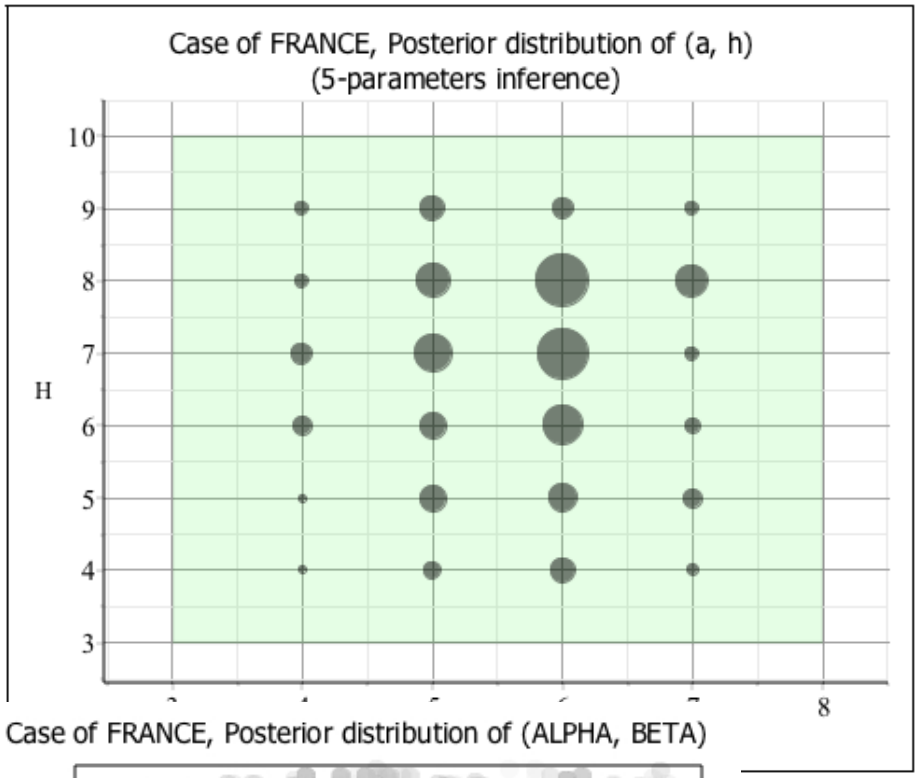
Marginal distribution of τ
(a kind of projection of Post(U) « in the direction τ » , https://en.wikipedia.org/wiki/Marginal_distribution)
This marginal distribution is all the more spread out as the interaction of τ with the four parameters **a**, **b**, α and β is important.

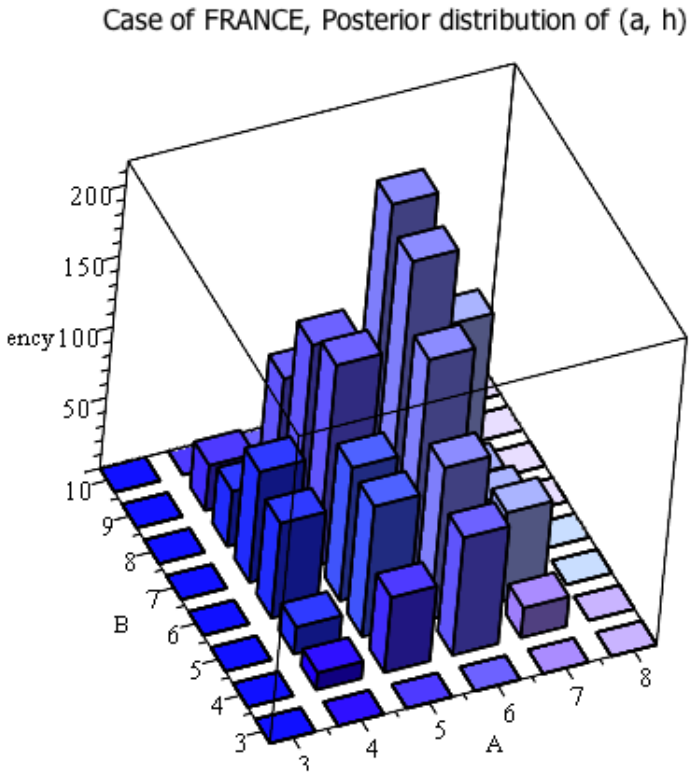
It is the uncertainty on these last four parameters that prevents a precise estimation of τ (thus the reason why they are termed « nuisance parameters » https://en.wikipedia.org/wiki/Nuisance_parameter).
Let me say that this term « nuisance » is misleading, and particularly here, for it suggests that the only quantity that matters is τ . But the survival delay **T** is just as important, and I'm sure some could say that **a** and **b** are extremely important and, why not, α and β too.

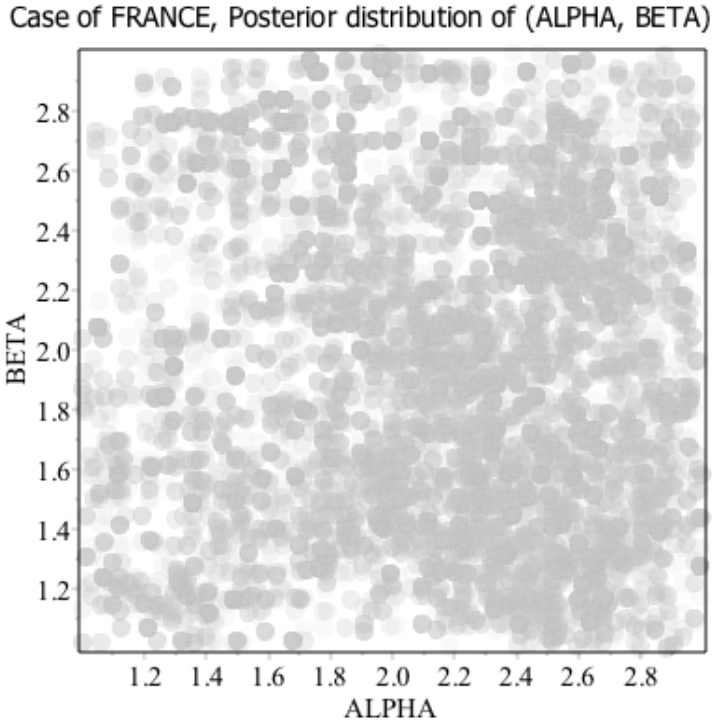
The histogram below clearly illustrates this, not nuisance, but « fuzzyfication » induced by the now unknown parameters **a**, **b**, α and β (simply compare this histogram to those given in the first treatment of the french data).



Marginal distribution of the couple (a, h)
The size of the grey discs centered at the integer coordinates i=3..8 and j=3..10 are proportional to the number of times the pair (a_i, h_j) has been met along the Markov chains equal to (i,j).... A representation in the form of a histogram is also provided.

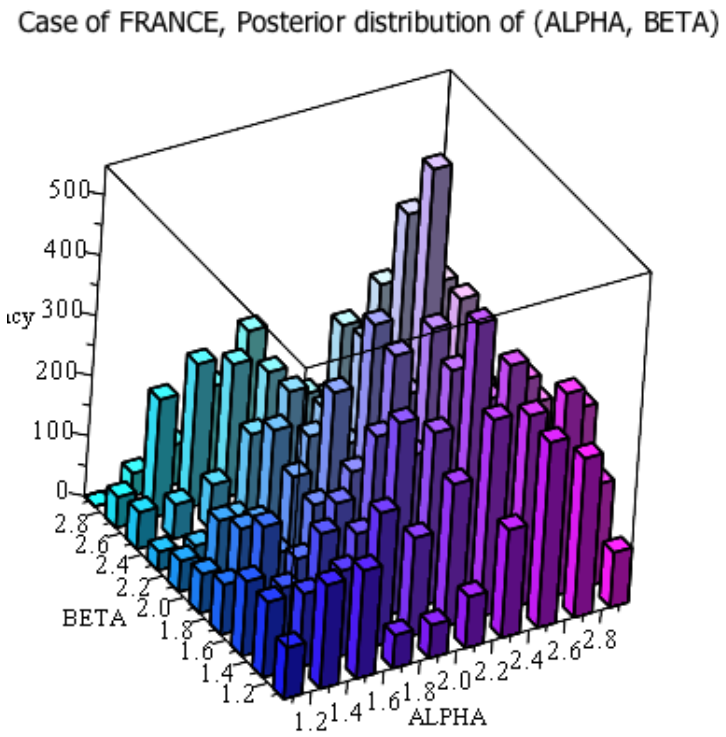






Marginal of the couple (α, β).

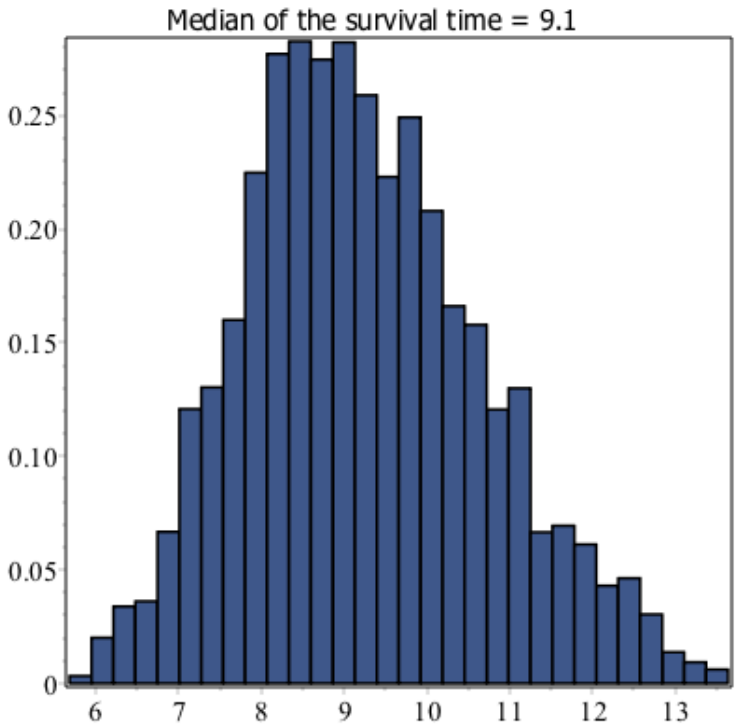
This map is interesting in that no clear structure seems to emerge from it beyond " α takes high values and β small values". This can be interpreted in two different ways : either the Bayesian Inference doesn't enable to learn much about these parameters, either their detailed joint distribution doesn't really matters.

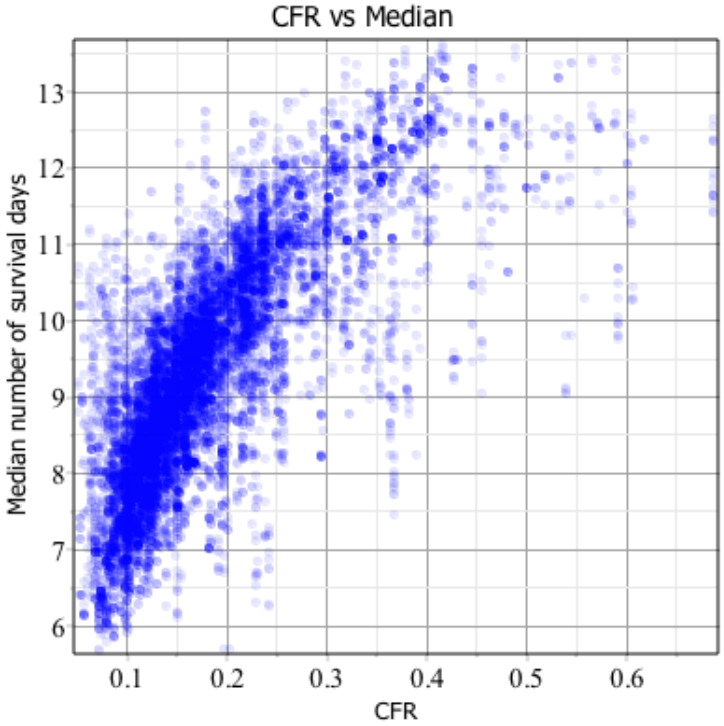


Posterior distribution of the median of T :

This median is rather concentrated around the value 9 (histogram on the left); the scatterplot on the right shows clearly how the the lack of knowledge of this median impacts the assessment of the mortality rate τ.

This is easily understandable: suppose that on day D there are 100 new cases detected, 200 on D+1 and 400 on D+2; if 50 people die on D+3 with a survival time exactly equal to 1 day, then the mortality rate is 50/400 or 1/8; but if this survival time is exactly 2 days the mortality rate rises to 50/200 = 1/4 and even to 1/2 if the survival time is exactly 3 days.





Posteriori (linear) correlation between components of U.

The correlation matrix below shows no significative (linear) correlations between **a**, **b**, α and β. The only clear (linear) correlation are between these latter and τ.

- When α increases, all the other parameters involved in the definition of **T** being kept at constant values, the probability density function of Beta(α, β) concentrates on its right and thus increases the value of the median of **T**. As we have saw above that the mortality rate is an increasing function of this median, the linear correlation between τ and α is expected to be positive.

- This same reasoning applied now to β (the probability density function of Beta(α, β) concentrates on its left) makes us to expect a negative (linear) correlation between τ and β.

	TAU	A	H	ALPHA	BETA
TAU	1.000	0.527	0.355	0.181	-0.316
A	0.527	1.000	-0.009	-0.091	-0.052
H	0.355	-0.009	1.000	-0.041	-0.035
ALPHA	0.181	-0.091	-0.041	1.000	0.002
BETA	-0.316	-0.052	-0.035	0.002	1.000