

# Beyesian Inference for Decision Making under COVID-19: A case study

## Abstract

Inference on circumstances where hypothesis have to be done with no data to substantiate them is tough. The following paper presents a case study on COVID-19 induced job loss and how the bayesian theory of inference sets clear ground rules for incorporating such hypothesis and the caveats of doing so with the Mathematical Statistics approach

## Introduction

The worldwide health crisis induced by the the " *SARS-COV 12* ", globally referred as COVID-19, mutated the classic decision making under uncertainty problem into a substancially more uncertain mess due to the lack of closeness of the 2020 pandemic to previous world plights. Under this particularly uncertain context Bayesian Theory of Inference bestows an adequate framework to incorporate such new precariousness.

Having a reliable ground where to stand becomes critical for policy makers or other high impact decision makers. The following paper presents a simple, yet non trivial, case study of Bayesian Inference to address the difficulty of predicting the job loss induced by the COVID-19 lock down measures. We also stress out some of the caveats of the classical approach from the Fisher and Neyman school of Mathematical statistics and contrast both approaches slanting outturns in the context of decision making.

Section **I** begins by providing the reader a concrete outlook of the problem and the data at hand to tackle it. Section **II** translates the problem into its concrete mathematical and statistical representation and comments on the frailty of classical solution. Section **III** goes about the same formulation under a simple *by-hand* bayesian alternative and introduces the concept of likelihood resilience, while section **IV** proposes a simpler solution using an alternative conjugate family. Section **V** presents a more sophisticated approach through Monte Carlo Markov Chain Stochastic simulation an finally section **VI** is a brief utterance on the importance of providing a well grounded inference for decision making processes.

## 1 Mexican Social Security Institute and lagged payrrrol notices

## 2 Estimating labor force decrease the “accepted way”

Having settled some common grounds we can formulate the former situation into statistical language.

Let  $N_t$  be the number of workers noticed on time or in advanced by some company to the Social Service Mexican Institute at some given month  $t$ . As mentioned before, it's of interest to predict the amounts of the labor force  $N_{t+1}$  for the successive month at the instances of the health crisis.

Which could be a sufficiently simple formulation of the problem? At first, the notation used and even the timely nature of the phenom may suggest some time series approach. However it's the objective of the authors to present a as simple as possible formulation for the model. Being that so, though it may look a bit naive at first, lets consider the following structural form for the lagged worker numbers at month  $t + 1$ , that being:

$$N_{t+1} = N_t \cdot (1 + \theta)$$

Notice that under the prior structural statement, the forecasting of  $N_{t+1}$  reduces to estimating  $\theta$  the *lagged augmentation parameter* of the number of workers between  $t$  &  $t + 1$ .

Yet even more naive and for the sake of simplicity lets assume  $x_i = (\frac{N_{t+1}}{N_t} - 1) \sim N(\theta, \sigma^2)$  and that  $\{x_i\}_{i=1}^N$  constitutes a collection of *i.i.d* random variables i.e. a *random sample*  $X_{(n)}$ . Going further  $\hat{\theta}$  and  $\hat{\sigma}^2$  would correspond to the MLE estimators, nothing new.

Although the normal parametrization is convenient for its simplicity, it should stand out to anyone that there is a major problem choosing the Gaussian family, this is so because  $\theta \in \Theta = [0, \infty]$  and under the normality assumption  $\Theta = [-\infty, \infty]$ . For now, our way around will be to truncate such  $N(\hat{\theta}, \hat{\sigma}^2)$  and redistribute tail density<sup>1</sup>.

In any context not involving such distressful times everything would work fine, in fact our Fisherian conclusion would be that with 95% confidence level  $\theta$  would lie within the **INSERTAR INTERVALO**<sup>2</sup>. However things start getting funny when incorporating unobserved hypothesis on the probable behavior of  $\theta_{COVID-19}$ . Lets explore some alternatives:

- a. Let Marcos be some well trained senior economist Ph.D on labor market dynamics, known for his work as a government adviser for the Mexican Ministry of Labor. As an experienced scientist and given the experience in similar (or not that similar countries) Marcos believes, in fact he is almost sure, that  $\theta_{COVID-19}$  will in fact be of around 5%. Still, the only way to reach 5% levels given the MLE framework would be arbitrary marking down *a posteriori*  $\hat{\theta}_{COVID-19}$  under the pretence of having strong evidence in favor of 5%.

Many colleagues would argue that this textbook like toy model doesn't even correspond to the ordered nature of observations or even the parametric space. However any other, more sophisticated, model would encounter the same inconvenience as there would be no input accounting for COVID-19 induced contraction at the time, not to mention the rise in model complexity and the need for computer power.

Although Marcos' opinion may be well founded and it could be right, it appears to be introduced into the inference machinery by coercing the likelihood's output giving it not voice nor vote. Additionally this *mark down* fashion of placing constraints on the data is not accompanied by the corresponding increase in standard deviation due to the rise in uncertainty of the actual  $\theta_{COVID-19}$  behavior.

- b. Katia, a fellow colleague of Marcos, suggests increasing  $\sigma_{MLE}^2$  by some amount to reflect the uncertainty. Marcos agrees with Katia, however the question that will inevitably arise is, by how much?.

Again, any given arbitrary quantity for increasing variance will be an exogenous input in the classical inference work flow. Although there may be certain rules of thumb from which the writers are currently unaware, in the context of decision making environments an inference which could structurally and consistently incorporate both Marcos' and Katia's insight would be preferred over case specific tips.

### 3 Incorporating reverend Bayes square table

Rescuing our "not-so-toy" example, an alternative to address some of the classical framework caveats is to proceed through bayesian methodss. We'll part from the same assumptions and the same structural construction from the model except, for now, we will assume variance to be known and equal to historical variance, additionally we assume a reference distribution for  $\theta$  given by  $N(\theta|\mu_0, \sigma_0^2)$  in order to take advantage

<sup>1</sup>This approach is nothing more but a conditional density.

<sup>2</sup>This interval length corresponds to the *predictive interval length* given by  $\frac{X_{n+1} - \bar{X}}{s_n \sqrt{1 + \frac{1}{n}}} \sim T^{n-1}$ . The interval is not symmetric because of the truncation.

of a conjugate scheme. Building the model this way, we get that the posterior predictive distribution for  $\theta_{t+1} = \theta_{COVID-19}$  is given by

$$f(X_{t+1}|X_{(n)}) \sim N(x|\mu_n, \sigma_n^2 + \sigma_x^2)$$

Where  $\mu_n = \sigma_n^2 \cdot (\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{x}}{\sigma_x^2})$  and  $\sigma_n^2 = \frac{1}{\frac{n}{\sigma_x^2} + \frac{1}{\sigma_0^2}}$ .

Commenting on the forgoing model, it could be said that we gain complexity at the cost of assuming a known variance, which in the authors' opinion is definitely not true. Nevertheless we gained so much more than a little mathematical complexity, that being a consistent way to introduce Marcos' expert knowledge into the inferential process as an endogenous input, rather than an exogenous distortion on the model, through  $\theta$ 's *a priori* distribution.

Following this line of thought, Marcos' expert knowledge is introduced in the form of a (highly) **informative prior**. Expertise is incorporated first by centering the future behaviour of the lagged number of employees around 5% and second by adjusting the level of "sureness" by declaring a very low standard deviation <sup>3</sup> lets say .001%. In fact choosing this particular characterization for  $\theta_{COVID-19}$ 's prior yields a prediction interval at 95% credibility level of **INSERTAR INTERVALO DE CREDIBILIDAD** around **INSERTAR ESTIMACIÓN PUNTUAL** <sup>4</sup>.

Even though this specific prior characterization is as arbitrary as any number chosen for a *mark-down procedure* the *a priori* and not *a posteriori* fashion of introducing this knowledge, plus the interaction between parameters through the bayesian inference frame work and the **likelihood resilience** <sup>5</sup> allow for a coherent manner to let hypothesis and data interact with each other in a parsimonious fashion within the walls of a model<sup>6</sup>.

From last paragraph its particularly worth discussing the term *likelihood resilience*. This formally defined as the **sensitivity of the posterior paramaters to extreme prior distributions** or in other words "how strong is the likelihood evidence of the data". Accounting for likelihood resilience is done by quantifying the relative change in the posterior distributions parameters from the original hypothesis by in the presence of extreme prior **FALTA PULIR ESA DEFINICIÓN CON MEDIDAS RELATIVAS ESPECÍFICAS EN TERMINOS DE MOVIMIENTOS EXTREMOS DE LA PRIOR PARA CIERTO TRHESHOLD**. Likelihood resilience can be though as an end-user type off insurance, in the way that reporting this metric along with the results prevents, at least in some degree, the "*lying with statistics phenom*". Further more its an objective measure of the robustness of the inference done in presence of an unobserved hypothesis. In Figure 1 we can appreciate the posterior's distribution behavior with different prior paramaters.

## FIGURE ONE

**DESCRIBIR COMPORTAMIENTO.** Likelihood resilience is an intrinsic characteristic arising from the way our inference was built. additionally in can be analytically traced and studied in conjugate schemes through the functional form of the estimators and numerically when applying MCMC methods.

Nonetheless many may be still be troubled by the known variance assumption, which is totally reasonable. Relaxing this premise is just a matter of a little more mathematical intricacy <sup>7</sup>. For a more realistic scenario we choose the Normal-Inverse Gamma family where the predictive distribution for  $X_{n+1}$  is:

$$f(X_{n+1}|X_{(n)}) \sim t_{2\alpha_n}(x|\mu_n, \frac{\beta_n(\kappa_n + 1)}{\alpha_n \kappa_n})$$

Where:

<sup>3</sup>Conversely with very high precision  $\tau = \frac{1}{\sigma^2}$

<sup>4</sup>After truncation

<sup>5</sup>Haro-Peniche 2020

<sup>6</sup>This also adresses concern **b** from section I

<sup>7</sup>All normal conjugacy cases are wel documented in: Murphy. *Conjugate analysis of the Gaussian distribution*. 2007. Avaiable here

- $\mu_n = \frac{\kappa_0 \mu_0 + n \bar{x}}{\kappa_0 + n}$
- $\beta_n = \beta_0 + \frac{1}{2} \sum_{i=1}^n (x_i - \bar{x})^2 + \frac{\kappa_0 n (\bar{x} - \mu_0)}{2(\kappa_0 + n)}$
- $\kappa_n = \kappa_0 + n$
- $\alpha_n = \alpha_0 + \frac{n}{2}$

We set prior hyperparameters to be **Insertar supuestos** to reflect a similar uncertainty as in the first case. As in the known variance case described some lines earlier, we need to truncate the posterior predictive distribution to adjust for  $\theta$ 's parametric space yielding the following 95% credibility interval of **INSERTAR INTERVALO**<sup>8</sup>.

Figure 2 displays a similar likelihood resilience tests for the posterior predictive distribution.

## FIGURA 2

### DESCRIPCIÓN FIGURA 2

Although analytic manipulation becomes a bit more messy than obtaining the MLE from the classical approach, it should be obvious by know the benefit of advocating for the later methodology. We gain a coherent inference substructure to include unobserved hypothesis, we get the likelihood resilience insurance for free and in the conjugate case we can trace the interaction between model parameters (which in the mark-down fashion we cannot).

## 4 The Exponential-Gamma system

We finally need to adress the parametric space issue as truncation may introduce some distortions of the interest parameter's behavior. To account for a more realistic representation of  $\Theta$ , in which case we could opt to characterize  $X_i$  as  $Exp(\theta)$  however the MLE will still have problem **a** and **b** as described in section I.

The bayesian alternative is to require the Exponential-Gamma conjugate family. Here  $X_i$  is distributed as before and  $\theta \sim \Gamma(\theta|\alpha_0, \beta_0)$  and the posterior predictive distribution is given by

$$f(X_{n+1}|X_{(n)}) \sim Pareto(x|\alpha_n, \beta_n)$$

Where

- $\alpha_n = \alpha_0 + n - 1$
- $\beta_n = \beta_0 + n \bar{x}$

Given this second conjugate scheme we account for the actual  $\Theta$ , in fact we gain a substantially simpler math than the normal conjugate case and again we acquire the likelihood resilience insurance for free. Under this new parametrization and setting  $\alpha_0 = \text{INSERTAR NÚMERO}$  and  $\beta_0 = \text{INSERTAR NÚMERO}$  we obtain a 95% credible interval of **INSERTAR INTERVALO** around **INSERTAR ESTIMACIÓN PUNTUAL**.

Again we can directly analyze likelihood resilience to different priors yielding figure 3.

## FIGURA 3

## 5 Going full MCMC bayesian

In spite of sections IV and V proposals we may still may want to try some other “out the box” solution, for example fitting a log-Normal distribution to the data. Although in this case the classical approach offers

<sup>8</sup>Truncation of posterior predictive t-distribution was done with r package *truncdist* v. 1.0-2.

an analytical solution through MLE, the output will lack, again, the robustness to incorporate COVID-19 diminished  $\theta$  hypothesis. On the other hand if we take the bayesian way we still have a conjugate case for the known variance case, however going beyond that hypothesis would require numerical methods, in particular, Markov Chain Monte Carlo stochastic simulation.

Although, at first, this may look as a downside, abandoning conjugate schemes to MCMC methods, if proceeding with caution, give us an amazing improvement in modeling flexibility.

For our last example let's consider  $X_i \sim \text{logNormal}(x|\theta, \sigma^2)$  with  $f(\theta) \sim \Gamma(\theta|\alpha, \beta)$  and  $f(\sigma^2) \sim \Gamma^{-1}(\sigma^2|a, b)$

**EL CASO LOGNORMAL ES UNA JALADA PERO ESTÁ MUY BUENO LO PUEDES ENCONTRAR DOCUMENTADO AQUÍ (SE HIZO CON R)** [https://projecteuclid.org/download/pdfview\\_1/euclid.ba/1354024469](https://projecteuclid.org/download/pdfview_1/euclid.ba/1354024469)

For the classical approach

On the other hand for the bayesian approach we would like to use

## 6 Wrapping up

Table 1 depicts  $\theta$ 's different values for in terms of job loss in Mexico with their correspondent regions.

**Table 1**

### **Comentar de ls diferencias entre estimadores**

Critical decisions for policy making, such as budget allocation for unemployment insurance or setting target employment levels arise from choosing a particular number from the table above, and any of these translates in a difference of hundreds or maybe millions in terms of cash. Moreover it's desirable to have as robust as possible quantity in which we can rely on.

The frequentist approach, as powerful as it is, may not be the best choice for a setting such the one we are facing now a days<sup>9</sup>. **CONLUIR**

---

<sup>9</sup>Besides many authors as Mendoza (**INSERTAR AÑO**) have documented many issues with the classical procedures such as failures in the construction of confidence intervals see