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

5/26/2020

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 particuarly uncertain context Bayesian Theory of Inference bestows an adecuate framework to incorporate such new precariousness.

In particular, 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 adress the difficulty of predicting the job loss induced by the COVID-19 lockdown measures. We also stress out some of the caveats of the classical aproach from the Fisher and Neyman school of Mathematical statistics and contrast both approaches slanting outyurns in the context of desicion making.

Section **I** begins by providing the reader a concrete outlook of the problem and the data at hando 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 while section **IV** proposes a more sophisticated solution and cautions the reader on the first asumptions. Finally section **V** is a brief utterance on the importance of providing a well grounded inference for decision making proceses.

Mexican Social Security Institute and lagged payrrol notices

Estimating labor force decrease the "accepted way"

Having setled 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 Insitute 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 succesive 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 posible formulation for the model so. Being that so, though it may look abit naive at first, lets consider the following structural form for the lagged worker numbers at $t + 1$ 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 θ 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 colection of $i.i.d$ random variables i.e. a *random sample* $X_{(\underline{n})}$. Proceding after the construction of this model $\hat{\theta}$ 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 because $\theta \epsilon \Theta = [0, \infty]$ 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 [1].

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

a. Let Marcos be some well trained senior economist Ph.D on labor market dynamics, well known for his work as a government advisor for the Mexican Ministry of Labor. As an experienced fellow 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 favour of 5%.

Many coleagues would argue that this textbook like toy model doesn´t even correspond to ordered nature of observations or even the parametral space. However any other, more sophisticated, model would encounter the same incovenience 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 giving the prior not voice nor vote. Additionaly this *mark down* fashion of placing constrains on the data is not accompanied by the corresponding increase in standard deviation due to the rise in uncertainty of the actual $\theta_{COVID-19}$ behaviour.

b. Katia, a fellow colleague of Marcos, suggests increasing σ^2_{MLE} by some amount to reflect the uncertainty. Marcos agrees with katia, however the question that will innevitably arise is, by how much?.

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

Incorporating reverend Bayes square table

Rescuing our "not-so-toy" example, an alternative to adress some of the classical framework caveats is to proceed through bayesian grounds. We'll part from the same assumptions and the same structural construction from the model except, for noew, we will asume variance to be known and equal to historical variance, additionally we assume a reference distribution for θ given by $N(\theta|\mu_0, \sigma_0^2)$ in orider to take advantage of a conjugacy scheme. Building the model this way, we get that the posterior predictive distribution for $\theta_{t+1} = \theta_{COVID-19}$ is given by

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

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

Commenting on the forgoing model, it could de said that we gain complexity at the cost os assuming a known variance, which in obviously not true. Nevertheless we gained so much more more than a little mathematical

[1] This approach is nothing more but a conditional density.

[2] This interval lenght corresponds to the *predictive interval lenght* given by $\frac{X_{n+1} - \overline{X}}{s_n\sqrt{1+\frac{1}{n}}} \sim T^{n-1}$

complexity, this being a consistent way to introduce Marco's expert knowledge into the inferential process as an endogenous input through θ 's *a priori* distribution.

Following this line of though, Marco's expert knowledge is introduced in the form of a (highly) **informative prior**. Expertise is incorporated first by centering the future bahaviour of the lagged number of ployees around 5% and second by adjusting the level of "surenees" by declaring a very low standard deviation [3] lets say .001%. In fact chossing this particular characterization for $\theta_{COVID-19}$'s prior yields an prediction credibility interval of **INSERTAR INTERVALO DE CREDIBILIDAD** around **INSERTAR ESTIMACIÓN PUNTUAL** [4].

Even though this specific prior characterization is as arbitrary as any number choosen for a *mark-down procedure* the *a prior* and not *a posteriori* fashion of introducing this knowledge, plus the interaction between paramters through the conjugacy scheme and the **likelihood resilience** [5] allow for a coherent manner to let hypothesis and data interact with each other parsimonious fashion within the walls of a model[6].

From last paragraph its particuarly worth discussing the term *likelihood resilience* from the known variance Normal-Normal conjugate family. Which is formally measured as the relative in change in the posterior distributions parameters in the presence of extreme prioir **FALTA PULIR ESA DEFINICIÓN CON MEDIDAS RELATIVAS ESPECÍFICAS EN TERMINOS DE MOVIMIENTOS EXTREMOS DE LA PRIOR PARA CIERTO TRHESHOLD**. In Figure 1 we can aprreciate the posteriors distribution behaviour with different prior paramters.

FIGURE ONE

DESCRIBIR COMPORTAMIENTO. In other words likelihood resilience in this model is not just an intrinsic characteristic that prevents arbitrary data coercion or the *lying with statistics* issue but a quantifiable measure of how robust the inference is against extreme hypotheis.

Nonetheless many may be still be troubled by the known variance assumption, which is in fact totally reasonable. Relaxing this premise is just a matter of a litter more mathematical intricacy [7]. For a more realistic scenario we choose the Normal-InverseGamma family where the predictive distribution for $\hat{\theta}_{COVID-19}$ is:

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

Where:

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

Additionally the posterior ma

[3] Conversely with very high precision $\tau = \frac{1}{\sigma^2}$

[4] After truncation

[5] Haro-Peniche 2020

[6] This also adresses concern **b** from section I

[7] All normal conjugacy cases are weel documented in **Insertar cita**