A Wizard Predicts New COVID-19 Cases

Gandalf Olórin Knowledge[Ω]

December 5, 2020

Abstract

I was given the number of daily new Norwegian COVID-19 cases for the period from February 21th to November 10th. It started off with 1 new case and on November 10th there were 599 new cases. I predicted, rounded to the nearest hundred , that there would be 700, 600, 700, 400, and 400 new cases the next five days. The next five days gave, rounded to the nearest hundred: 700, 600, 700, 400, and 400 new cases. This note explains the magic. [1]

Keywords: **conditional probability; likelihood; time series;**

Figure 1 shows the daily number of new COVID-19 cases in Norway. Based on this I was asked to predict the daily numbers for November 11th to 15th. The actual daily numbers are now known, and are presented together with my predictions (and two greeningly inferior predictions) in Figure 2. The aim of this note is to explain the magic in elementary terms. Note, however, that a true understanding can only be obtained after years of study at a major school of magic - and these are predicted to disappear in the near future. So, hurry on!

It is clear that the future numbers of COVID-19 occurrences are uncertain. My predictions are based on conversations with fellow wizards Bernoulli (1713, *Ars conjectandi)*, Pearson (1920, correlation), Kolmogorov (1933, conditional expectation), Doob (1953, stochastic processes), Fisher (1973, likelihood), and Engle (1982, conditional variance model). [2] It turns out, for the given data, that including a conditional variance model outperforms a purely conditional mean model as shown in Figure 2. Explanations of some of the main ideas in the conversation with my fellow wizards follow below.

According to Bernoulli (1713, ch 2 in part 4), the art of conjecture is:

[1][Ω] *Beware! The One Space, Ω, rules them all!* Disclaimer: This is NOT research on COVID-19. Transcribed from communications in norønt by the humble servant G. Taraldsen.

[2] Robert Fry Engle III, together with Clive Granger, received the 2003 Nobel Memorial Prize in Economic Sciences for methods of analyzing economic time series with time-varying volatility.

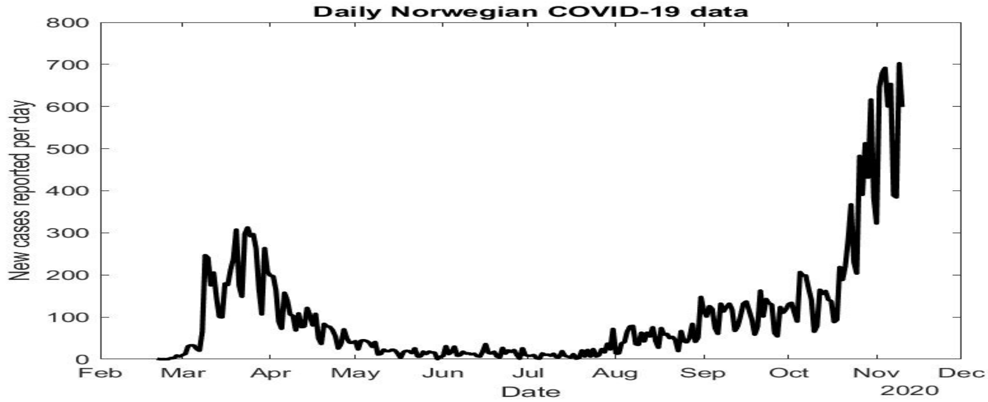

Figure 1: Daily numbers $y_{-7}, y_{-6}, \ldots, y_{256}$ of new reported cases of COVID-19 in Norway from February 21th to November 10th 2020.

> *The art of measuring, as precisely as possible, probabilities of things, with the goal that we would be able always to choose or follow in our judgments and actions that course, which will have been determined to be better, more satisfactory, safer or more advantageous.*

I will now apply this art to our problem. Let $A = (U = u)$ be the event corresponding to the observed numbers $u = (y_{-7}, y_{-6}, \ldots, y_{256})$ of COVID-19 occurrences on days from February 21th to November 10th. Let $B = (V = v)$ be the event corresponding to the five unknown [3] numbers $v = (y_{257}, \ldots, y_{261})$ of COVID-19 occurrences from November 11th to November 15th. The uncertainty of the unknown event B, when A is known, is given by the conditional probability

$$\mathrm{P}(B \mid A) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(A)} \tag{1}$$

The interpretation of this conditional probability is, as explained by the Bernoulli (1713, golden theorem, part 4) law of large numbers: The theoretical limit of the relative number of success, but restricted to outcomes in A.

How can the conditional probability $\mathrm{P}(B \mid A)$ be used to predict the COVID-19 cases on November 11th-15th? Let $f(u, v) = \mathrm{P}(A \cap B)$ and let $f(v \mid u) = \mathrm{P}(B \mid A)$. It follows then that f is a probability density in the sense that $\sum_{u,v} f(u, v) = 1$. It also follows that $f(v \mid u)$ is a conditional density in the sense that $\sum_v f(v \mid u) = 1$, and $f(u, v) = f(v \mid u) f(u)$ from equation (1).

How can the conditional density $f(v \mid u)$ be used for prediction? More generally: How can the probability distribution of a random quantity R be used to predict

[3] Unknown then, on November 10th, or unknown if you have not seen the five numbers.

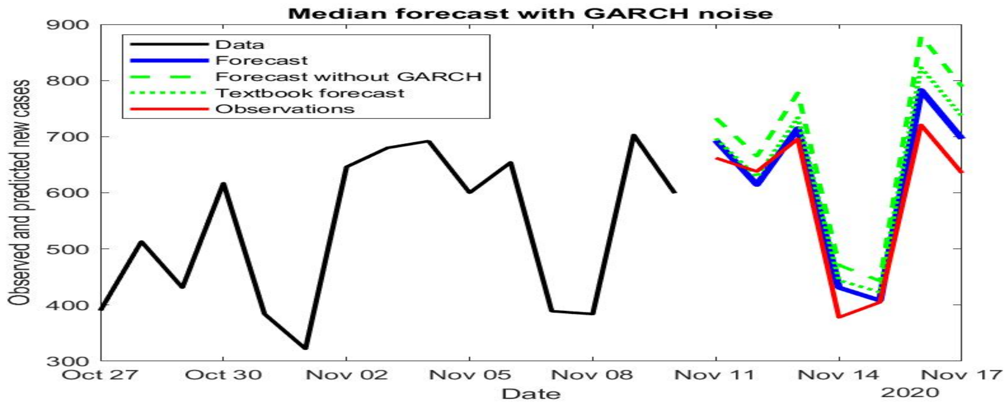

Figure 2: Observed and predicted cases of COVID-19.

its value? The textbook answer is the expectation $\mathrm{E}(R)$. Then, for our problem, the textbook answer is the conditional expectation

$$\mathrm{E}^u(V) = \sum_v v f(v \,|\, u) \tag{2}$$

This is not the best choice in this case as demonstrated in Figure 2. A better choice is the component-wise conditional median from the conditional density $f(v \,|\, u)$. The forecasts in Figure 2 have been computed simply as the empirical median and empirical mean of 1001 samples from the conditional density $f(v \,|\, u)$. What can we learn from this? Students of magic should not only learn about the expected value, but also learn about the median, and its relatives.

But wait! The conditional density $f(v \,|\, u)$ is unknown since the density f is unknown. This is true, so the sampling giving Figure 2 has been done based on an estimate of f. Beware! This is the point in our tale where true magic appears.

Elementary estimation of a density f can be done starting with for instance a histogram based on repeated observations from the density. A slightly improved method is sometimes given by introducing a parametric model $f^\theta(y)$ for the density, followed by estimating the model parameter θ by setting a suitable number of moments equal to the corresponding empirical moments. These methods give nothing, or very little, if only one observation from the density f is given. In our case, we do not even have 1 single observation from the density! All that is given is the part u of 1 observation $y = (u, v)$. True magic is needed!

Fisher (1973) comes in, after Bernoulli (1713), and solves the problem: Use the model density f that maximizes the likelihood of the actual observation. The likelihood is, in our case, given by

$$L(\theta) = \prod_{t=-7}^{256} f(u_t \,|\, u_{t-1}, \ldots, u_{-7}) \tag{3}$$

with the understanding that f depends on θ. For our problem it turns out that a particular model where θ is given by four numbers performs well. The estimate of θ used for the simulations giving Figure 2 was obtained by numerical optimization of the likelihood in equation (3).

How did I come up with a parametric model for our problem? A first convenient simplification is to consider the transformation $Y_t = \text{floor}(\exp(S_t))$. It follows that a model for S_t gives a model for Y_t, so I seek a model for S_t. One motivation behind the transformation is that S_t can take all real numbers as its values. Another motivation is that a symmetric distribution for S_t gives a skewed distribution for Y_t. Further motivation exist, but I will remain silent on this.

Equation (3), and it's version with s_t replacing u_t, motivates to consider a recursion equation where s_t is determined by it's past values $s_{t-1}, s_{t-2}, \ldots$, and a white noise source term. Assuming the white noise is Gaussian it turns out that Pearson (1920) correlation together with some helpful hints from Doob (1953) give a model with a reasonable prediction as shown in green in Figure 2.

Investigating the previous resulting noise term shows, however, that the white noise cannot be Gaussian. The estimated noise is white, but squaring the noise gives a correlated noise sequence. Now! Engle (1982) comes in and saves the day. He showed that this phenomena can be modeled by a conditional variance model. The resulting model, after some pondering and analysis (doi:10.13140/RG.2.2.23540.37769), is given by three equations: $(1-B)(1-B^7)S_t = (1-0.35(4)B)(1-0.32(5)B^7)Z_t$, $Z_t = \sqrt{h_t}W_t$, and $(1-0.90(1)B)h_t = 0.001 + 0.10(1)BZ_t^2$. Here B is the back-shift operator, $BX_t = X_{t-1}$, and $\ldots, W_{-1}, W_0, W_1, \ldots$ is a doubly infinite random sample from the standard normal distribution.

The notation $0.35(4) = 0.35 \pm 0.04$ indicates, in accordance with ISO (1995), a standard uncertainty equal to 0.04 with a corresponding approximate expanded uncertainty of 0.08 for a confidence level of 95%. A similar interpretation holds for the three remaining estimates of the components of θ. All students, and magicians, be aware! It is useful to have a standard for uncertainty!! Just as it is useful to have standard units for length, mass, and similar: All is provided by the International Organization for Standardization (ISO), and it's partners.

But! Kolmogorov (1933)? Yes, he has been there all the time, but hidden as in most texts on magic. There is, One Space, Ω, that rules it all. The axioms of the probability space Ω are presented in elementary texts on magic, but most often in a faulty way. Many say, for instance, that an event by definition is the same as a subset of Ω. Some magicians play the trick of using this erroneous definition when teaching younger students of magic. Shame on them! Thou shalt not lie! It is wrong. Every subset is not an event. Just as every function is not continuous or infinitely differentiable.

Solving the above given three equations involve limits of infinite sums of random variables. Proof of this requires the foundation given by the axioms of Kolmogorov (1933), or some generalization thereof. The same holds for the definition of the expectation $\mathrm{E}(R)$, and countless other concepts. Without the axioms, the magic disappears, bewildering dominates, and all that is left is endless handwaving when 'paradoxes' appear.

Tyche, the goddess of chance, also known as Fortuna in the positive minded Roman culture, plays the lead figure in this tale. Tyche selected the One Outcome ω in the One Space Ω before the appearance of time. As a goddess she did this only once - of course. By this she rules all events such as floods, droughts, frosts, or even in politics, when no cause can be discovered. This was described already by the Greek historian Polybius, but I provide here the version formulated more recently by Kolmogorov (1933).

In our example, we are given the value $u = U(\omega)$, but ω remains unknown. In fact, the One Space Ω, that rules it all, will also always remain unknown. How can I then, at November 10th, say anything about the value $v = V(\omega)$ for the days to follow? This has already been explained! A summary of the arguments are given by

(i) Axioms ensure existence of all probabilities $\mathrm{P}(A) = \mathrm{P}\{\omega \mid U(\omega) = u\} \ldots$

(ii) The likelihood gives an estimate of the model.

(iii) Prediction is based on the conditional probabilities $\mathrm{P}(B \mid A)$.

The above argument can be enhanced if a Bayesian prior distribution for $\theta = \Theta(\omega)$ can be motivated. Alternatively, the fiducial arguments of Fisher (1973) can be used. In either case, the last two steps are then modified accordingly. Tyche, the goddess of chance, rules, of course, also over the choice of model θ. As noted, there is One Space, Ω, that rules them all. Beware! Tyche is Ω.

The methods used for estimation and forecast are standard, and can most possibly be improved. It is, however, a challenge, for You, to get predictions that outperforms the ones given in Figure 2. Hints are possibly given by the finishing list of comments focused on the given prediction problem.

1. It's hard to be a Nissemann, and also a Bayesian. But interesting and fun!

2. How can a prior be chosen - without interpretation of the model parameters?

3. Objective Bayes? Fiducial inference? Confidence distribution?

4. Focus parameter? Loss function? Choice of model?

5. Can machine learning[4] outperform the classics?

Funding for research in true magic is, sadly, scarce and decreasing in Norway. The government decided to put a 'university' on every hill, and spread the already lacking funding accordingly. There are many hills of different and dark origin - and also so in Norway. True schools of magic will not be there in the future. The financing structure ensures that this is universally so. Formerly true schools of magic are transformed to the consultancy firms and research institutions with which they are competing with in 'research' proposals. Product development, with deadlines and clearly formulated work packages, is the golden standard.

The teaching and education are measured in a similar way, but now guided by popularity quests. Some magicians, possibly guided by this misguiding popularity quest, choose to demonstrate 'magic' for students by pointing line by line on computer scripts and the resulting graphics. This disease has gone even further in pre-magic schools: The exams are divided in two parts, where one part is point-and-click repetition of what the pre-teachers have demonstrated. The student majority are often very happy with this. How could they know better? They are students, and hence lacking in knowledge. This lack is exactly why they are students! The politicians, possibly also somewhat lacking in knowledge, are superhappy with this: The 'university' is then using the modern tools of computer science. Hurra, Hurray! The magic, which is in the subject matter, is lost.

References

Bernoulli, J. (1713). *Ars Conjectandi (The Art of Conjecturing, Translated by Edith Sylla 2005).* Johns Hopkins Univ Press.

Doob, J. L. (1953). *Stochastic Processes.* Wiley.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica 50*(4), 987–1007.

Fisher, R. A. (1973). *Statistical Methods and Scientific Inference.* Hafner press.

ISO (1995). Guide to the Expression of Uncertainty in Measurement. *International Organisation for Standardization.*

Kolmogorov, A. (1933). *Foundations of the Theory of Probability.* Chelsea (1956).

Pearson, K. (1920). Notes on the history of correlation. *Biometrika 13*(1), 25–45.

[4] Read: Modern computer intensive statistical methods. Beware! The classical methods are also developing. Time has not stopped.