

# Automated Forecasting

Junshi

February 16, 2022

## Abstract

Your abstract.

## 1 Introduction

Your introduction goes here! Simply start writing your document and use the Recompile button to view the updated PDF preview. Examples of commonly used commands and features are listed below, to help you get started.

Once you're familiar with the editor, you can find various project setting in the Overleaf menu, accessed via the button in the very top left of the editor. To view tutorials, user guides, and further documentation, please visit our [help library](#), or head to our plans page to [choose your plan](#).

## 2 Simple Model for Election Fraud

Here is the Metaculus [link](#) for this question. I now try to use a very simple and inaccurate model, and do a forecast with data accessible on [Google Search](#), keyword set as "election fraud".

### 2.1 Model Setting

Our problem of interest is "Probability of greater than 20% people believe election rigged". Let's define  $b$  as a random variable corresponding to the individual's belief in a rigged election.  $b = 1$  if one believes this firmly, and  $b = 0$  otherwise. Then  $p = \mathbf{E}(b)$  is the proportion of people firmly believe in the fraud election. So  $\mathbf{P}(p > 0.2)$  is what we cares about. This  $p$  also follows a normal distribution(I adopt the Bayesian approach).

As for Google Search, denoted as  $I$ , has a relationship with  $b$  and  $p$ . Intuitively, all those with  $b = 1$  with search this topic definitely while the others will pay attention to this with some probability(random variable  $s$ ). I assume  $s$  to follow a normal distribution in the month when everyone is talking about this and it becomes a constant 0.

To summarize, Statistical Model is

$$\begin{aligned} b &\stackrel{iid}{\sim} \mathbf{Bino}(1, p) \\ \text{Prior} : p &\stackrel{iid}{\sim} \mathbf{N}(1/2, 1/9) \\ I &\sim \mathbf{N}(p + s(1 - p), 1/9) \\ s &\sim \mathbf{N}(1/2, 1/9) \mathbf{1}\{T \leq 1\} \end{aligned}$$

Why 1/9? Only to simulate a truncated normal.

### 2.2 Result

In the month when election was really trendy(Nov 2021), the posterior distribution of  $p$  is  $p \stackrel{iid}{\sim} \mathbf{N}(\frac{2}{5}\hat{I} + \frac{9}{20}, 4/45)$ ,  $\hat{I} = 0.28$ . To estimate the believers of a rigged election in the following year(except Nov 2021),  $\hat{I} = 0.05$ ,  $p \stackrel{iid}{\sim} \mathbf{N}(\frac{1}{2}\hat{I} + \frac{1}{4}, 1/18)$ .

The final estimation of  $\mathbf{P}(p > 0.2)$  is then 0.76 and 0.59 respectively. According to [Reuters](#), in Nov 2020, 68% of Republicans said they were concerned that the election was "rigged," while only 16% of

Democrats and one-third of independents were similarly worried. However, in 2021 Nov, [report](#) shows that only 31% of the interviewees are still suspicious.

For reference, the Metaculus Community gave a almost consistent 80% confidence for this question and the question was resolved positively.

## 2.3 Comments

The model is extremely simplified. Better work can be done to have some hyper-parameters for  $p$  and more data should be used in prediction since the final outcome in this model is largely relied on our prior believe.

# 3 Simple Model for Cultured Meat

Here is the Metaculus [link](#) for this question. I now try to use a very simple and inaccurate model, and do a forecast based on the information provided by [Wikipedia](#), keyword set as "cultured meat".

## 3.1 Methodology

In this question, we are asked to predict the probability that cultured meat will be served in a restaurant before 2021/1/1. We think of this question abstractly as "what is the possibility that a scientific invention can be brought to the public within certain time period", so as to find our reference class, "scientific inventions", biological scientific interventions or inventions regarded as a novel consumption good.

Further unveiling the requirements of a successful invention, we may break it down into small stages of breakthroughs and assume that, on average an invention will be successful after  $n_{break}$  times of small breakthroughs have occurred. For example, in the case of cultured meat, the first patent in 1991 and the first public trial on TV in 2013 should be counted. To get a rough list of all breakthroughs, most events included in History section on Wikipedia are counted (This of course, is just a rough list, since we does not even have a rigorous definition of breakthroughs. However, there are some potential ways to do this, i.e. list their search times on google, ask an expert).

## 3.2 Model Setting

For this breakthrough model, we can readily adopt a regular poisson process model, where the probability of a breakthrough happening in any time unit is a constant  $\lambda$ . A bayesian's approach is again used here to estimate and update our estimation of  $\lambda$  over time.  $N(t)$  is defined as the total number of breakthroughs at time.

To summarize, Statistical Model is

$$\begin{aligned} \text{Prior : } \lambda &\sim \mathbf{Gamma}(\alpha, \beta) \\ N(t) &\sim \mathbf{Poi}(\lambda t) \\ \lambda | N(t) &\sim \mathbf{Gamma}(\lambda + N(t), \beta + t) \end{aligned}$$

Hyperparameter  $\alpha$  and  $\beta$  are chosen such that the prior belief imitate a uniform distribution.

## 3.3 Result

If we acknowledge the time of breakthroughs are recognized to be 1991,2001,2003,2008,2013,2017,2018,2019,2019,2020,2020 (out the one when cultured meat had its debut at [Singapore](#)), the model predictions the following probability at different combinations of  $n_{break}$  and  $t_p$  (the end of each year) which means the time of forecasting.

We can see that when  $n_{break}$  is set to be 9, 10 or 11, the prediction agrees best with the community.

$n_{break}$	2016	2017	2018	2019	2020
7	0.82	1.00	1.00	1.00	1.00
8	0.57	0.76	1.00	1.00	1.00
9	0.35	0.49	0.69	1.00	1.00
10	0.20	0.29	0.43	1.00	1.00
11	0.12	0.18	0.28	0.64	1.00
12	0.09	0.14	0.21	0.42	1.00
13	0.07	0.12	0.19	0.33	0.66
14	0.07	0.11	0.18	0.30	0.55
15	0.07	0.11	0.18	0.30	0.52

Table 1: Predictions based on Poisson Model

<i>Type</i>	2016	2017	2018	2019	2020
Community	0.38	0.80	0.72	?	?
Metaculus	0.22	0.48	0.50	?	?

Table 2: Predictions based on Poisson Model

### 3.4 Comments

Better work can be done to estimate hyperparameters(the ones we set manually, such as  $n_{break}$ ). It would be possible if  $n_{break}$  itself is random(follow a normal distribution maybe, as suggested by Will in the last call on Feb10th). Mostly important, a stricter definition or model for breakthroughs should be discussed. The full timeline is not accessible on Metaculus, so how well the model truly performs is hard to say at this point.

## References