Affects of government interventions on the spread of novel coronavirus

George Nadareishvili

Tbilisi State University

Seminar Mathematical Modelling of Epidemics

April 9, 2021

Bayesian philosophy

Steve is very shy and withdrawn, invariably helpful but with very little interest in people or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail.

სტივი ძალიან მორცხვი და ჩაკეტილია, ყოველთვის დასახმარებლად მზად მყოფი, მაგრამ რეალურ სამყაროსა თუ ადამიანებში ნაკლებ დაინტერესებული. მშვიდსა და პედანტს, წესრიგისა და დეტალების მიმართ განსაკუთრებული მიზიდულობა ახასიათებს.

Is Steve more likely to be a librarian or a farmer?

სტივის ბიბლიოთეკარობა უფრო მაღალალბათურია თუ ფერმერობა?

Bayes' rule

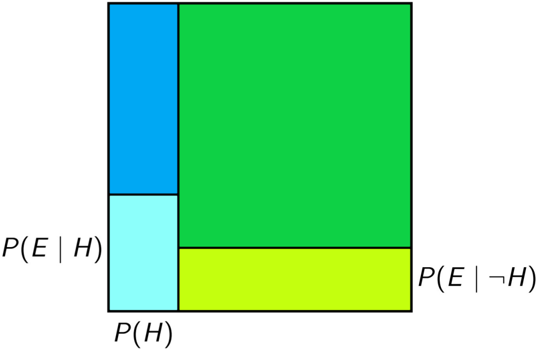
- This example is by Daniel Kahneman and Amos Tversky. People tend to ignore the overall proportion of farmers and librarians in the general population.
- The correct framework to think about this problem is Bayes' theorem.

$$P(H \mid E) = \frac{P(E \mid H) \cdot P(H)}{P(E)}$$

This helps to determine the probability of the hypothesis H (Steve is a librarian) given the evidence E (Steve is very shy and withdrawn. . .) is true.

- This is one of the most tattooed formulas in mathematics.
- To illustrate what the theorem says, it helps to draw a diagram.

Bayes' rule illustrated



- H Hypothesis, "Steve is a librarian"
- E Evidence, "Steve is very shy and withdrawn..."
- $\neg H$ Opposite of our hypothesis, "Steve is a farmer"

$$P(H \mid E) = \frac{P(E \mid H)P(H)}{P(E \mid H)P(H) + P(E \mid \neg H)P(\neg H)} = \frac{P(E \mid H)P(H)}{P(E)}$$

Terminology

In order to proceed, we have to learn some terminology from Bayesian statistics.

$$P(H \mid E) = \frac{P(E \mid H) \cdot P(H)}{P(E)}$$

$P(H)$ is called a prior

$P(E \mid H)$ is called a likelihood

$P(H \mid E)$ is called a posterior

- So, prior and likelihood are proportional to the posterior.
- Motto of Bayesian statistics: evidence, coming in, should update our prior beliefs, contrary to establishing completely new ones.
- This brings us to Bayesian inference.

Bayesian inference

In a nutshell: given a number of data samples (observations) that follow some distribution, what is the estimate of parameters at play?

Example

Say we are given 50 data samples $x_1, x_2, \ldots, x_{50}$. We know that this data samples are following some exponential distribution $a\mathrm{e}^{-ax}$. What is the estimation of a parameter a? There are at least two major estimation strategies

- Point estimation. A single best approximation

$$a = \Big(\frac{x_1 + x_2 + \cdots + x_{50}}{50}\Big)^{-1}.$$

  Prior beliefs are not taken into account.

- Interval estimation (or Bayesian inference). We get out a distribution of a, and we can also take prior beliefs into account.

Example in detail

First, lets generate 50 random data points (observations) that follow some exponential distribution.

```
n = 50 #the number of data points
a = RR.random_element(1,6) #random real number from interval [1,6] used as a parameter
import numpy # import library for random sampling
X = numpy.random.exponential(1/a, n) #generate "n" data points from exponential distribution with parameter "a"
```

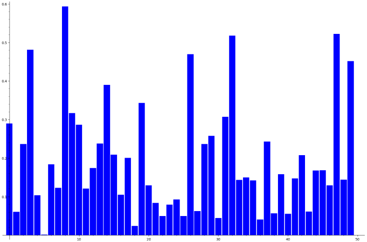


Figure: Data points

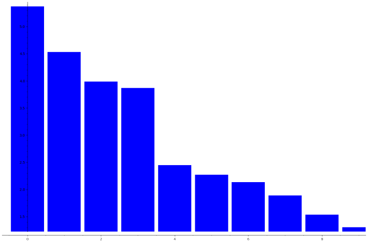


Figure: Histogram

Example in detail

For the sake of it, we find out what the point estimation of the parameter a. We also print the real value of a.

```
est = n/sum([X[i] for i in range(n)]) # point estimation of a parameter "a"
print('real value =',a,',','estimation =', est)
```

```
real value = 5.22953523145752 , estimation = 5.078818805838824
```

Step 1

We assume prior distribution of parameter a. Here, say we think it is distributed according to a normal distribution f with mean 4 and standard deviation 1. This is my prior.

Step 2

Generate random a's according to the prior distribution.

In our example, let us generate one thousand a's from our normal distribution f.

Example in detail

```
k = 1000 #number of random a's
mean = 4
div = 1
A = numpy.random.normal(loc = mean, scale = div, size= k) #generate "k" data
    points from normal distribution
```

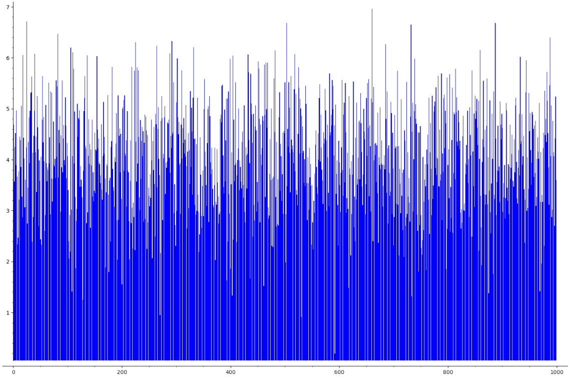


Figure: 1000 samples of a parameter from f

Example in detail

Step 3

Compute likelihood times prior $L(a_i)$ for each generated a_i.

$$
\begin{aligned}
L(a_i) &= \text{probability of observing } x_1, \ldots, x_{50} \text{ (if } a = a_i) \text{ and } a = a_i \\
&= P(x_1, x_2, \ldots, x_{50} \mid a = a_i) \times P(a = a_i) \\
&= P(x_1 \mid a = a_i) \times \cdots \times P(x_{50} \mid a = a_i) \times P(a = a_i) \\
&= (a_i \mathrm{e}^{-a_i x_1}) \mathrm{d}x \times \cdots \times (a_i \mathrm{e}^{-a_i x_{50}}) \mathrm{d}x \times f(a_i) \mathrm{d}x
\end{aligned}
$$

```
for s in range(k):
    for u in range(n):
        P[s,u] = A[s]*e^(-A[s]*X[u])*(1/n) #P[i,j] = a_i * e^(a_ix_j)*dx
```

```
f(x) = (2*pi*div^2)^(-1/2)*e^(-(x-mean)^2/2*div^2) #prior distribution function
for i in range(k):
    Pri[i] = (f(A[i]))*(1/k) #prior probablities for each a_i
```

```
    L[i] = prod([P[i,u] for u in range(n)])*Pri[i] #liklihood␣
↪P(x|a=a_i)*P(a=a_i)
```

Example in detail

Step 4

Draw the posterior distribution of a parameter a.

By Bayes' rule

$$P(a = a_i \mid x_1, x_2, \ldots, x_{50}) = \frac{L(a_i)}{\sum_{i=0}^{1000} L(i)}$$

```
S = sum([L[i] for i in range(k)])
for i in range(k):
    Pos[i] = L[i]/S  # posterior
```

Plotting the posterior distribution in blue, prior distribution in red and real value of a in green, we get

Example in detail

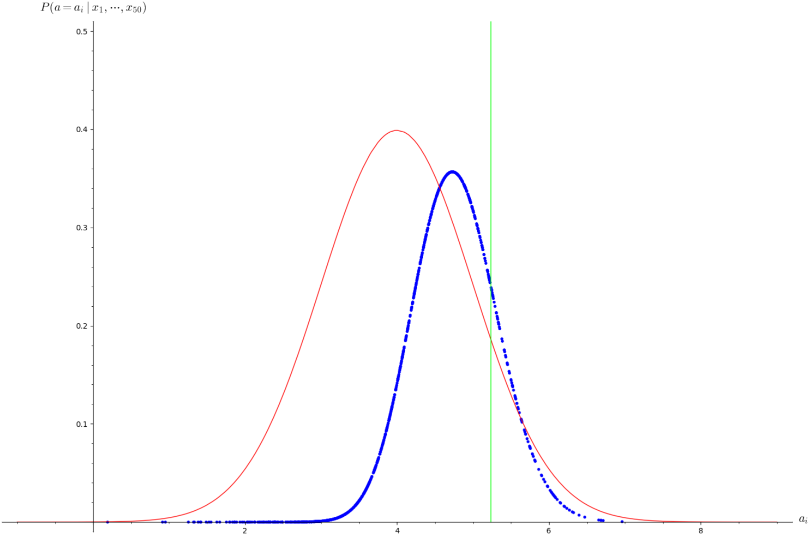


Figure: Prior, posterior and real value of a.

Example in detail

We can also draw the confidence interval.

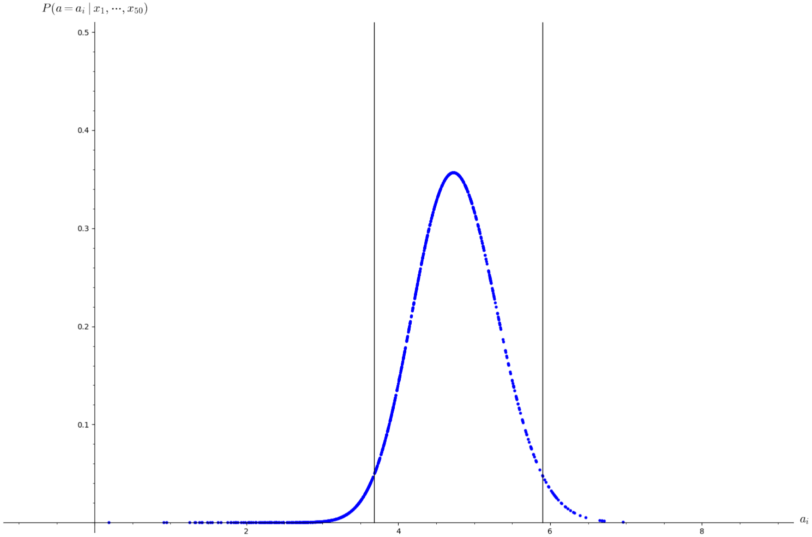


Figure: 95% confidence interval

Example in detail

How the number of observed data influence the estimation?

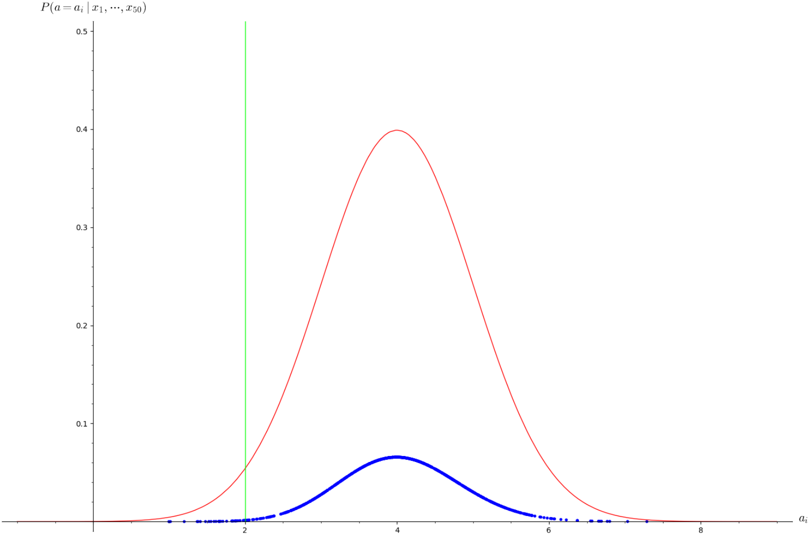


Figure: 10 data samples

Source

From now on, everything we say follows

Jonas Dehning, Johannes Zierenberg, F. Paul Spitzner, Michael Wibral, Joao Pinheiro Neto, Michael Wilczek, and Viola Priesemann.
Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions.
Science 369 (2020), no. 6500.

Bayesian inference for epidemiological parameters

We perform Bayesian inference for the epidemiological parameters of an SIR model using MCMC sampling (lecture by Ana Dolidze).

The central parameters are (there are others)

- the spreading rate λ,
- the recovery rate μ,
- the reporting delay D,
- the number of initially infected people I_0.

Informative priors are chosen based on available knowledge for λ, μ, and D, and uninformative priors for the remaining parameters.

The informative priors are intentionally kept as broad as possible so that the data would constrain the parameters.

Recall the SIR model

Disease spreads at rate λ from the infected population compartment I to the susceptible compartment S, and that the infected population compartment is removed R at rate μ.

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\lambda\frac{SI}{N}, \qquad \frac{\mathrm{d}I}{\mathrm{d}t} = \lambda\frac{SI}{N} - \mu I, \qquad \frac{\mathrm{d}R}{\mathrm{d}t} = \mu I.$$

N is population size. During the initial phase: $S \approx N \gg I$, so

$S/N \approx 1$. So, differential equation reduces to

$$\frac{\mathrm{d}I}{\mathrm{d}t} = (\lambda - \mu)I, \text{ solved by } I(t) = I_0\mathrm{e}^{(\lambda-\mu)t}.$$

Discrete SIR model

Dataset discrete in time ($\Delta t = 1$ day) $\implies$ solve equations in discrete time step ($\mathrm{d}I/\mathrm{d}t \approx \Delta I/\Delta t$)

$$
\begin{aligned}
S_t - S_{t-1} &= -\lambda \Delta t \frac{S_{t-1}}{N} I_{t-1} && =: -I_t^{\text{new}} \\
R_t - R_{t-1} &= \mu \Delta t I_{t-1} && =: R_t^{\text{new}} \\
I_t - I_{t-1} &= \left(\lambda \frac{S_{t-1}}{N} - \mu\right) \Delta t I_{t-1} && = I_t^{\text{new}} - R_t^{\text{new}}
\end{aligned}
$$

I_t models the number of all (currently) active infected people.

I_t^{new} is the number of new infections that will eventually be reported.

Add change points in spreading rate

- Assume the spreading rate λ_i, $i = 1, \ldots, n$, may change at certain time points t_i from λ_{i-1} to λ_i.
- Assume change happens linearly over time window of Δt days.
- Thereby, we account for mitigation measures, which were implemented by governments step by step.
- The parameters t_i, Δt_i, and λ_i are added to the parameter set of the model above.
- Differential equations are augmented by the time-varying λ_i.

Add reporting delay and weekly reporting modulation

- We explicitly include a reporting delay D between new infections I_t^{new} and newly reported cases C_t.
- We also include a weekly modulation to account for lower case reports around the weekend, which subsequently accumulate during the week.
- To model these, we put

$$C_t = I_{t-D}^{\text{new}}\Big(1 - f(t)\Big), \text{ where}$$
$$f(t) = (1 - f_w) \cdot \left(1 - \left|\sin\left(\frac{\pi}{7}t - \frac{1}{2}\Phi_w\right)\right|\right).$$

- Parameters f_w and Φ_w are also constrained by data.

Estimating model parameters

Denote by $\hat{C}_t$ the real-word data of reported cases.

- For likelihood $P(\hat{C}_t \mid \theta)$, assume the distribution

$$P(\hat{C}_t \mid \theta) \sim \mathrm{StudentT}_{\nu=4}\Big(\mathrm{mean} = C_t(\theta), \\ \mathrm{width} = \sigma\sqrt{C_t(\theta)}\Big).$$

- The scale factor σ is another parameter to be estimated.
- Student's t distribution because it resembles normal distribution around the mean but features heavy tails, which make the sampling more robust with respect to extreme values and thus reporting noise.

Estimating model parameters

Summing up, the complete set of model parameters is

$$\theta = \{\lambda_i, t_i, \Delta t_i, \mu, D, \sigma, I_0, f_w, \Phi_w\}$$

We estimate θ using Bayesian inference.

- The posterior is approximated by normal distributions ignoring correlations between parameters.
- Informative priors on initial model parameters are chosen where data allows.
- Uninformative priors otherwise.

Forecasting

- For the forecast, we take all samples from the inference step and continue time integration according to different forecast scenarios.
- The overall procedure yields an ensemble of forecasts as opposed to a single forecast that would be solely based on one set of (previously optimized) parameters.

Example: Germany

Three change points to the initial spreading rate λ_0, motivated by German government interventions.

- The first change point ($\lambda_0 \to \lambda_1$) expected around 9 March 2020 (t_1) as a result of the official recommendations to cancel large events.
- A second change point ($\lambda_1 \to \lambda_2$) expected around 16 March 2020 (t_2), when schools and many stores were closed.
- A third change point ($\lambda_2 \to \lambda_3$) is expected around 23 March 2020 (t_3), when all nonessential stores were closed and a contact ban was enacted.

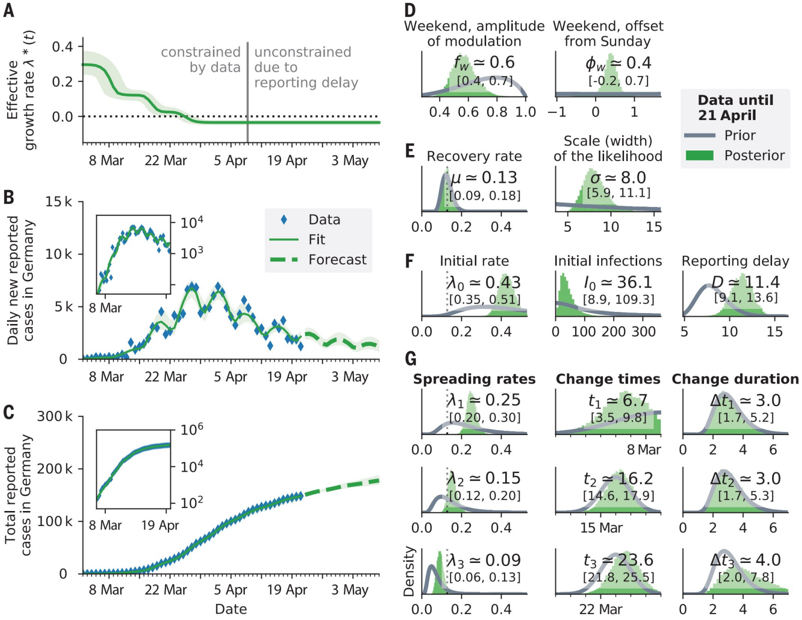
Assume that each set of interventions leads to a reduction by $\sim$50% of λ at a change point.
Assume wide uncertainty of reduction. In principle, even an increase at the change point would be possible after inference.

Germany: Priors

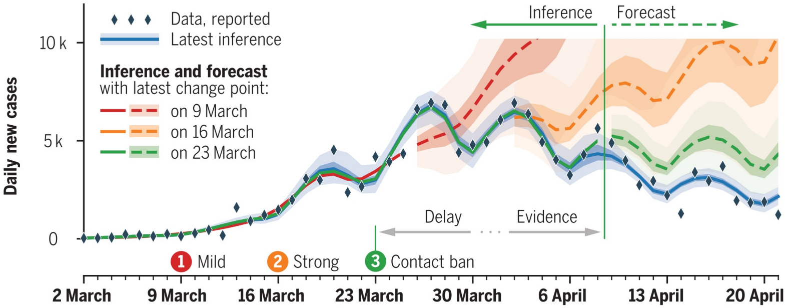
Exact priors (chosen because reasons) for model parameters are given below.

Parameter	Variable	Prior distribution
Change points	t_1	Normal(2020/03/09, 3)
	t_2	Normal(2020/03/16, 1)
	t_3	Normal(2020/03/23, 1)
Change duration	Δt_i	LogNormal[log(3), 0.3]
Spreading rates	λ_0	LogNormal[log(0.4), 0.5]
	λ_1	LogNormal[log(0.2), 0.5]
	λ_2	LogNormal[log(1/8), 0.5]
	λ_3	LogNormal[log(1/16), 0.5]
Recovery rate	μ	LogNormal[log(1/8), 0.2]
Reporting delay	D	LogNormal[log(8), 0.2]
Weekly modulation amplitude	f_w	Beta(mean = 0.7, std = 0.17)
Weekly modulation phase	Φ_w	VonMises(mean = 0, κ = 0.01) (nearly flat)
Initially infected	I_0	HalfCauchy(100)
Scale factor	σ	HalfCauchy(10)

Germany: Results



Germany: Results



Bayesian inference of SIR model parameters from daily new cases of COVID-19 enables us to assess the impact of interventions. In Germany, three interventions (mild social distancing, strong social distancing, and contact ban) were enacted consecutively (circles). Colored lines depict the inferred models that include the impact of one, two, or three interventions (red, orange, or green, respectively, with individual data cutoff) or all available data until 21 April 2020 (blue). Forecasts (dashed lines) show how case numbers would have developed without the effects of the subsequent change points. Note the delay between intervention and first possible inference of parameters caused by the reporting delay and the necessary accumulation of evidence (gray arrows). Shaded areas indicate 50% and 95% Bayesian credible intervals.

Germany: Results

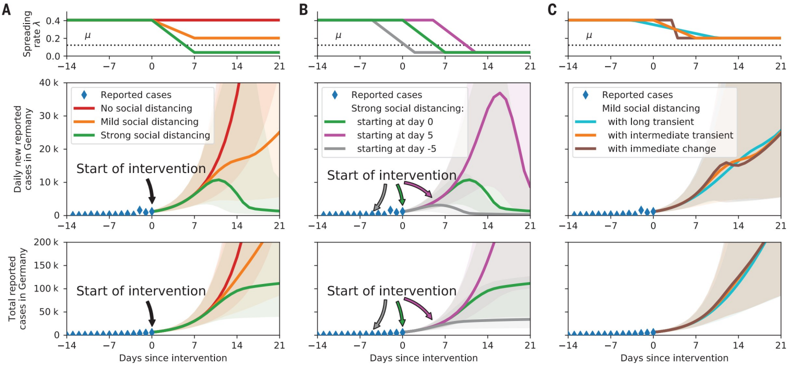


Fig. 2. The timing and effectiveness of interventions strongly affect future COVID-19 cases. (**A**) We assume three different scenarios for interventions starting on 16 March 2020: (i, red) no social distancing, (ii, orange) mild social distancing, or (iii, green) strict social distancing. (**B**) Delaying the restrictions has a major impact on case numbers: strict restrictions starting on 16 March 2020 (green), 5 days later (magenta), or 5 days earlier (gray). (**C**) Comparison of the time span over which interventions ramp up to full effect. For all ramps that are centered around the same day, the resulting case numbers are fairly similar; however, a sudden change of the spreading rate can cause a temporary decrease of daily new cases (although $\lambda > \mu$ at all times; brown).

Thank you!