EECS Research Students Symposium - 2020

COVID-19 Infection Rate Estimation

12th July, 2020

Navin Kashyap
Manjunath Krishnapur

Vinay Kumar B.R.
Arvind Rameshwar

Jnaneshwar B.S.
Anand Jerry George

Problem Description

GOAL: Estimate the fraction of infected individuals in the population

$$\mathcal{I} := \frac{\text{Number of infected individuals}}{\text{Total size of the population}}$$

Problem Description

GOAL: Estimate the fraction of infected individuals in the population

$$\mathcal{I} := \frac{\text{Number of infected individuals}}{\text{Total size of the population}}$$

Known: Daily data of tests, available from hospitals/testing centers

- Number of tests conducted
- Number of +ve tests

Problem Description

GOAL: Estimate the fraction of infected individuals in the population

$$\mathcal{I} := \frac{\text{Number of infected individuals}}{\text{Total size of the population}}$$

Known: Daily data of tests, available from hospitals/testing centers

- Number of tests conducted
- Number of +ve tests

Why?

- Extent of spread
- Design testing strategies
- Appropriate medical facilities

Problem Description

GOAL: Estimate the fraction of infected individuals in the population

$$\mathcal{I} := \frac{\text{Number of infected individuals}}{\text{Total size of the population}}$$

Known: Daily data of tests, available from hospitals/testing centers

- Number of tests conducted
- Number of +ve tests

Why?

- Extent of spread
- Design testing strategies
- Appropriate medical facilities

NOTE: We want to estimate $\mathcal{I}$, not predict it.

Formulation

Given: The time series

$P(t) :=$ Number of hospitalized individuals who test positive on day t.

To find:

$\mathcal{I}(t) =$ Number of infected individuals on day t

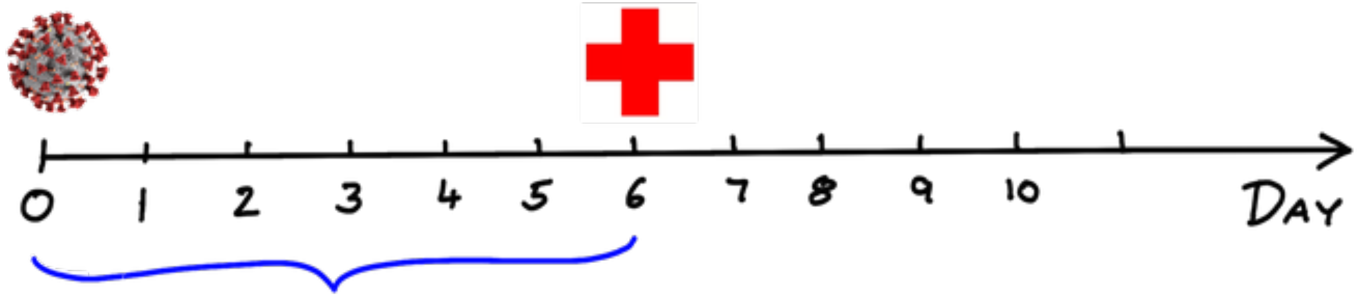
Formulation

Given: The time series

$P(t)$:= Number of hospitalized individuals who test positive on day t.

To find:

$\mathcal{I}(t)$ = Number of infected individuals on day t



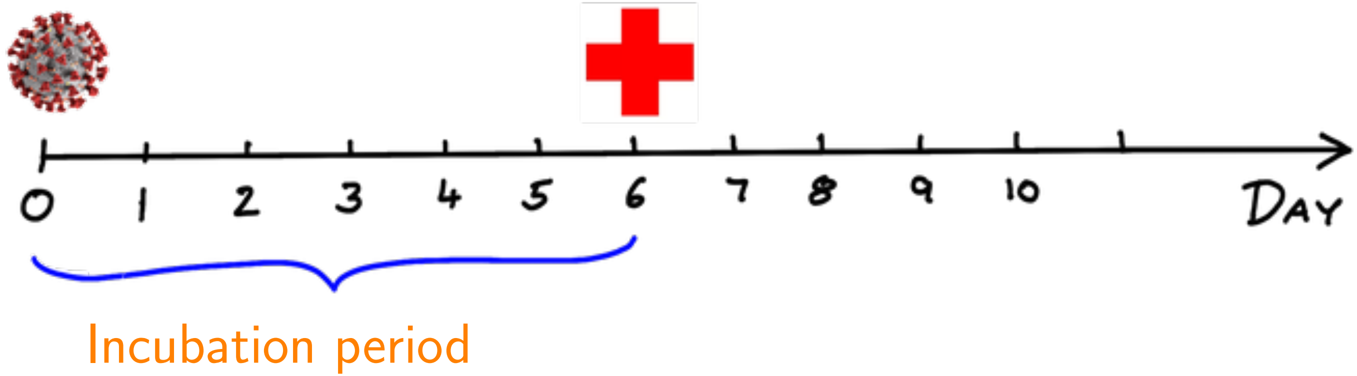
Formulation

Given: The time series

$P(t) :=$ Number of hospitalized individuals who test positive on day t.

To find:

$\mathcal{I}(t) =$ Number of infected individuals on day t



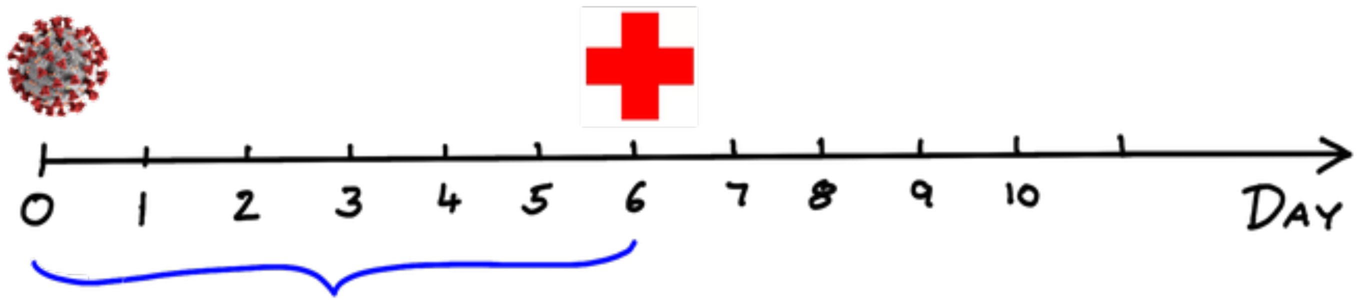
Formulation

Given: The time series

$P(t) :=$ Number of hospitalized individuals who test positive on day t.

To find:

$\mathcal{I}(t) =$ Number of infected individuals on day t



Incubation period

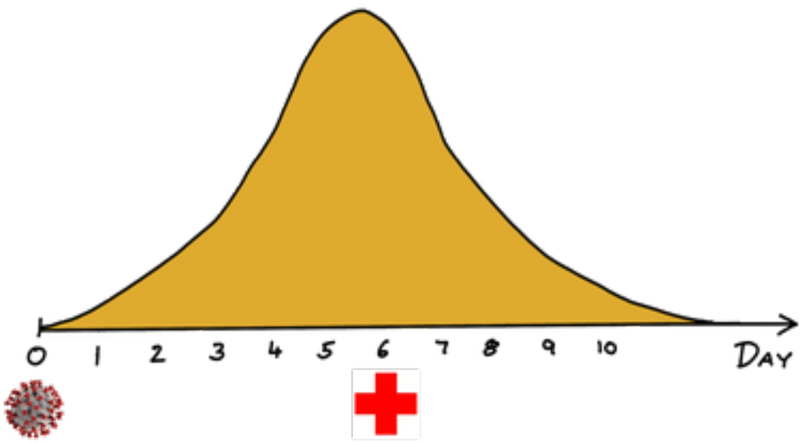
Assumptions:

- Onset of symptoms and tested positive on the same day
- No missed detections

Incubation Period Distribution

$\mathbf{q} = (q_1, q_2, \cdots, q_K)$

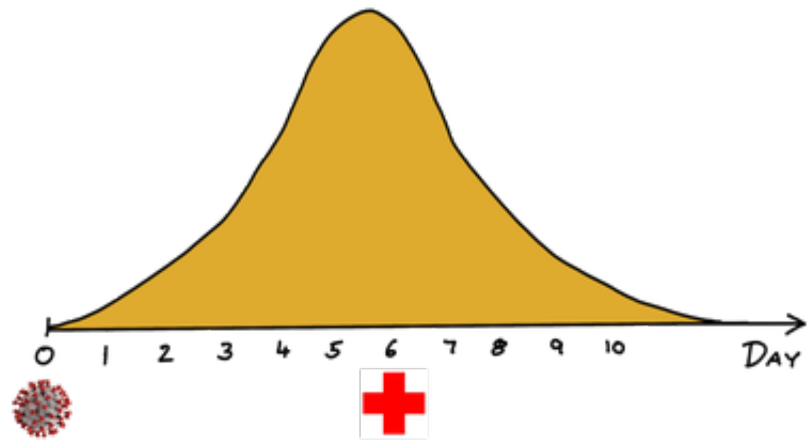
$q_k =$ Probability that an infected individual shows symptoms on day k after getting infected.



Incubation Period Distribution

$\mathbf{q}=(q_1, q_2, \cdots, q_K)$

q_k= Probability that an infected individual shows symptoms on day k after getting infected.



Remarks:

- Log-normal distribution with median of 5.1 days and mean 5.5 days [a]
- Max. number of days before the onset of symptoms $K = 14$

[a] Lauer et al. (2020), The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. Annals of Internal Medicine, 10th March 2020. https://annals.org/aim/fullarticle/2762808/incubation-period-coronavirus-disease-2019-covid-19-from-publicly-reported

Formulation contd...

On day t,

$$P(t) = \mathcal{I}(t-K) \times q_K + \mathcal{I}(t-K+1) \times q_{K-1} + \cdots + \mathcal{I}(t-1) \times q_1$$

Formulation contd...

On day t,

$$P(t) = \mathcal{I}(t-K) \times q_K + \mathcal{I}(t-K+1) \times q_{K-1} + \cdots + \mathcal{I}(t-1) \times q_1$$

Asymptomatic fraction

$q_\infty =$ Fraction of individuals who are infected but do not show any symptoms

- Not known exactly
- Unclear whether infectious or not

Formulation contd...

On day t,

$$P(t) = \mathcal{I}(t-K) \times q_K + \mathcal{I}(t-K+1) \times q_{K-1} + \cdots + \mathcal{I}(t-1) \times q_1$$

Asymptomatic fraction

q_∞ = Fraction of individuals who are infected but do not show any symptoms

- Not known exactly
- Unclear whether infectious or not

$$P(t) = (1 - q_\infty)[\mathcal{I}(t-K) \times q_K + \mathcal{I}(t-K+1) \times q_{K-1} + \cdots$$

$$\cdots + \mathcal{I}(t-1) \times q_1]$$

Estimator

$$P(t) = (1-q_\infty)[\mathcal{I}(t-K)\times q_K + \mathcal{I}(t-K+1)\times q_{K-1} + \cdots$$

$$\cdots + \mathcal{I}(t-1) \times q_1]$$

Estimator

$$P(t) = (1 - q_\infty)[\mathcal{I}(t-K) \times q_K + \mathcal{I}(t-K+1) \times q_{K-1} + \cdots$$

$$\cdots + \mathcal{I}(t-1) \times q_1]$$

For $t = K+1, \cdots, m$ (current day), write in matrix form

$$\mathbf{P} = Q \times \mathcal{I}$$

Estimator

$$P(t) = (1-q_\infty)[\mathcal{I}(t-K)\times q_K + \mathcal{I}(t-K+1)\times q_{K-1} + \cdots$$

$$\cdots + \mathcal{I}(t-1)\times q_1]$$

For $t = K+1, \cdots, m$ (current day), write in matrix form

$$\mathbf{P} = Q \times \mathcal{I}$$

$\mathbf{P}$: $[P(K+1)\ P(K+2)\cdots P(m)]^T$

Q: Toeplitz matrix

$\mathcal{I}$: $[\mathcal{I}(1)\ \mathcal{I}(2)\ \cdots\ \mathcal{I}(m-1)]^T$

Estimator

$$P(t) = (1-q_\infty)[\mathcal{I}(t-K)\times q_K + \mathcal{I}(t-K+1)\times q_{K-1} + \cdots$$

$$\cdots + \mathcal{I}(t-1)\times q_1]$$

For $t = K+1, \cdots, m$ (current day), write in matrix form

$$\mathbf{P} = Q \times \mathcal{I}$$

$$[P(K+1)\ P(K+2)\cdots P(m)]^T \qquad [\mathcal{I}(1)\ \mathcal{I}(2)\ \cdots\ \mathcal{I}(m-1)]^T$$

Toeplitz matrix

Estimator for $\mathcal{I}$

$$\hat{\mathcal{I}} = \arg\min_{\mathcal{I}\geq 0} ||\mathbf{P} - Q\mathcal{I}||^2 + \lambda \times \sum_{t=2}^{m} (\mathcal{I}(t) - \mathcal{I}(t-1))^2$$

How good is it?

$$\hat{\mathcal{I}} = \arg\min_{\mathcal{I} \geq 0} ||\mathbf{P} - Q\mathcal{I}||^2 + \lambda \times \sum_{t=2}^{m} (\mathcal{I}(t) - \mathcal{I}(t-1))^2$$

Generative model

- Population $= 100000$
- Initial infected $= 100$
- $R_0 = 1.1$
- $\mathcal{I}(k) \propto 100 \times (R_0)^k$
- $m = 40$ days
- $\mathbf{q}_{gen} \sim Unif([0, 14])$

This generates $P(t)$ and $\mathcal{I}(t)$

For estimation

- $\mathbf{q}_{est} \sim Unif([0, 14])$
- $\lambda = 0.1$

How good is it?

$$\hat{\mathcal{I}} = \arg\min_{\mathcal{I} \geq 0} ||\mathbf{P} - Q\mathcal{I}||^2 + \lambda \times \sum_{t=2}^{m} (\mathcal{I}(t) - \mathcal{I}(t-1))^2$$

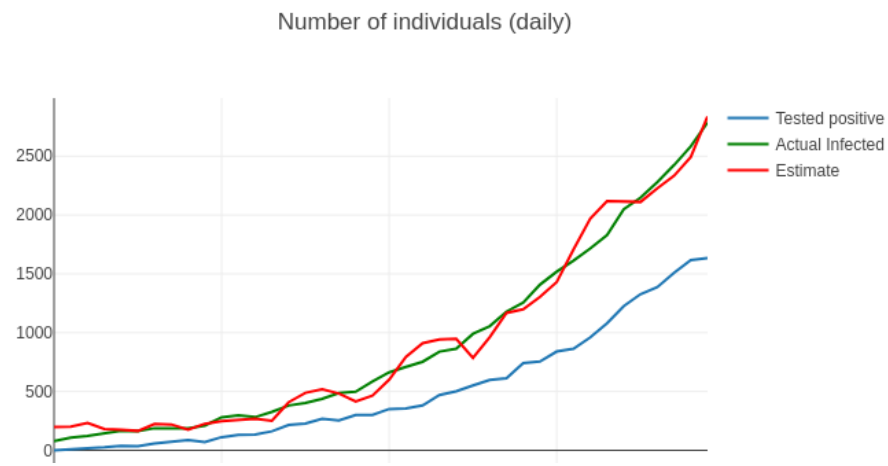
Generative model

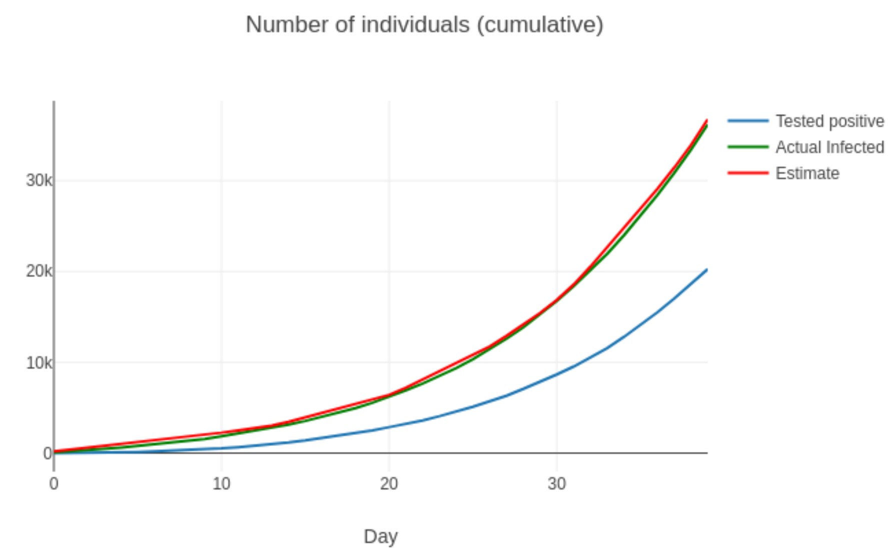
- Population $= 100000$
- Initial infected $= 100$
- $R_0 = 1.1$
- $\mathcal{I}(k) \propto 100 \times (R_0)^k$
- $m = 40$ days
- $\mathbf{q}_{gen} \sim Unif([0, 14])$

This generates $P(t)$ and $\mathcal{I}(t)$

For estimation

- $\mathbf{q}_{est} \sim Unif([0, 14])$
- $\lambda = 0.1$





Robustness towards q

$$\hat{\mathcal{I}} = \arg\min_{\mathcal{I} \geq 0} ||\mathbf{P} - Q\mathcal{I}||^2 + \lambda \times \sum_{t=2}^{m} (\mathcal{I}(t) - \mathcal{I}(t-1))^2$$

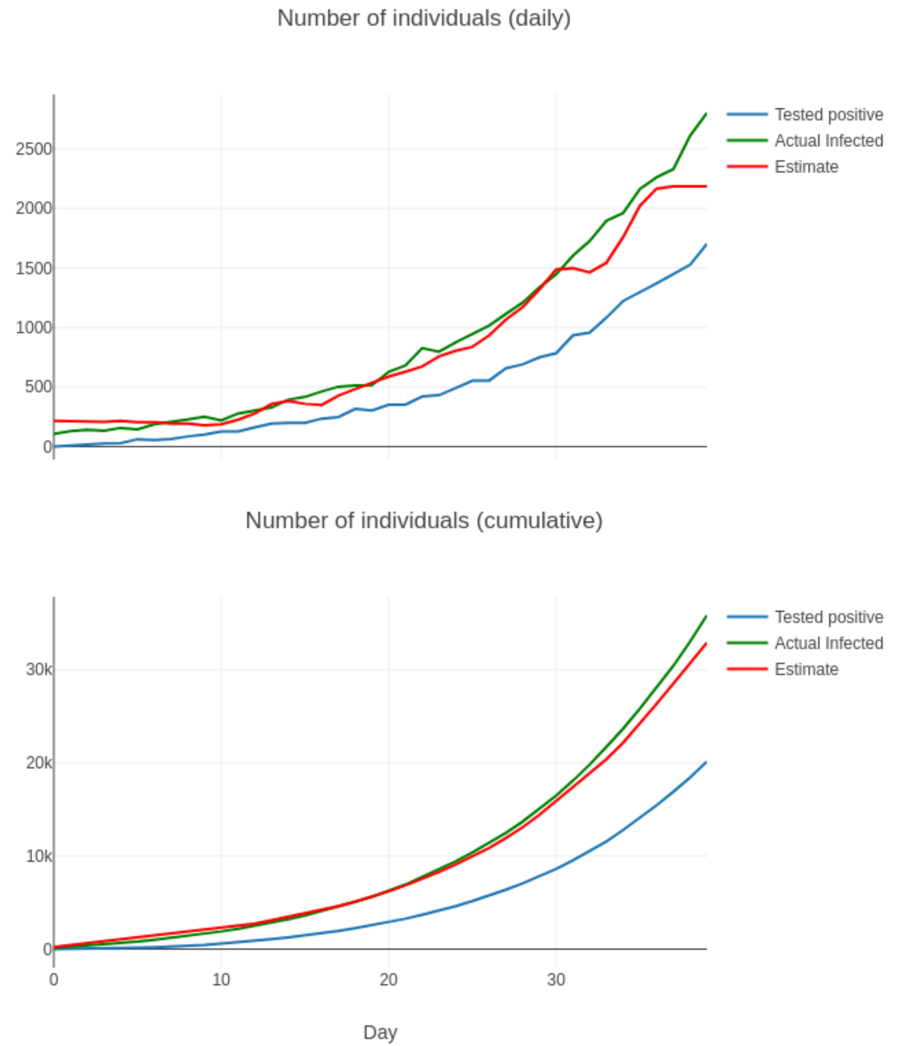
Generative model

- Population $= 100000$
- Initial infected $= 100$
- $R_0 = 1.1$
- $\mathcal{I}(k) \propto 100 \times (R_0)^k$
- $m = 40$ days
- $\mathbf{q}_{gen} \sim Unif([0, 14])$

This generates $P(t)$ and $\mathcal{I}(t)$

For estimation

- $\mathbf{q}_{est} \sim LogNormal$
- Median$=5.1$, Mean$=5.5$
- $\lambda = 0.1$



Robustness towards support of q

$$\hat{\mathcal{I}} = \arg\min_{\mathcal{I} \geq 0} ||\mathbf{P} - Q\mathcal{I}||^2 + \lambda \times \sum_{t=2}^{m} (\mathcal{I}(t) - \mathcal{I}(t-1))^2$$

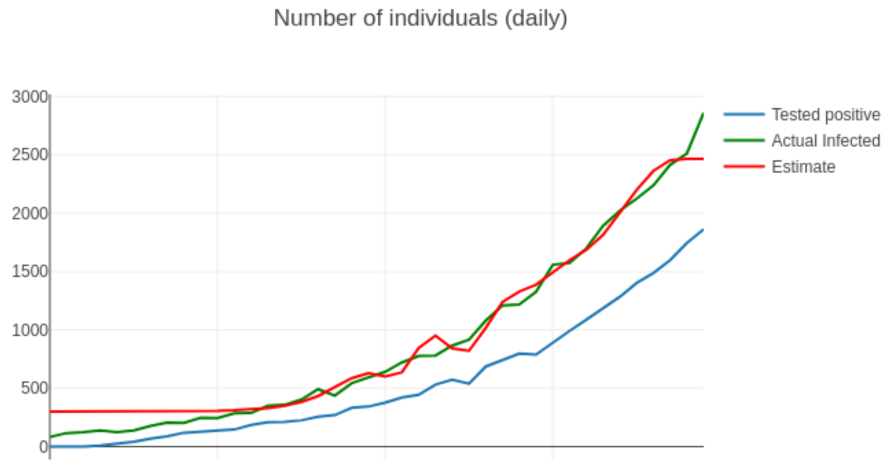
Generative model

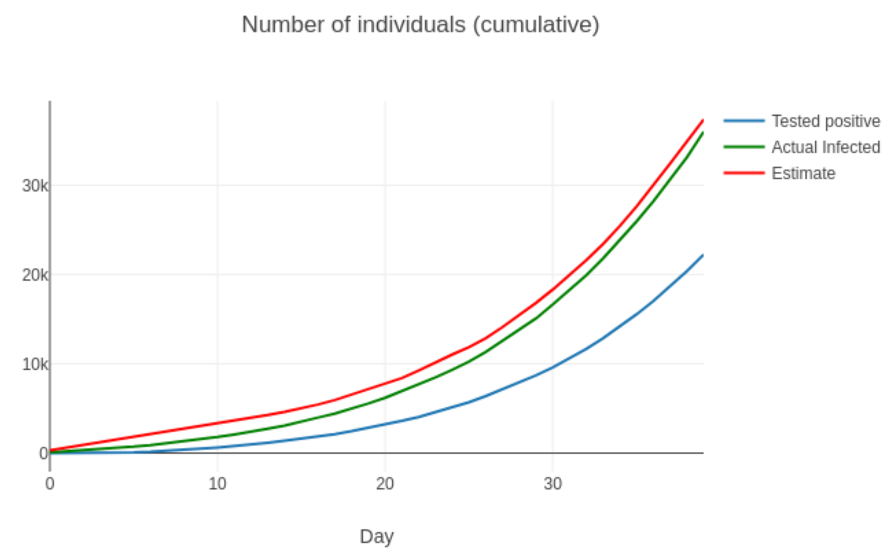
- Population $= 100000$
- Initial infected $= 100$
- $R_0 = 1.1$
- $\mathcal{I}(k) \propto 100 \times (R_0)^k$
- $m = 40$ days
- $\mathbf{q}_{gen} \sim LogNormal$
- $K = 10$

This generates $P(t)$ and $\mathcal{I}(t)$

For estimation

- $\mathbf{q}_{est} \sim LogNormal$
- $K = 20$
- $\lambda = 0.1$





On Actual data

https://covid19-prev-est.herokuapp.com/

On Actual data

https://covid19-prev-est.herokuapp.com/

Caveats

- Tested positive data includes contacts and other high risk individuals
- Asymptomatic fraction is not known

Thank you