In OR_Sim_SIR_MacPan.R we have implemented a fitting mechanism with MacPan2 for a fixed time frame, i.e.

- We simulate the real data from $t = 0$ and the observed real data is from a time period between t_{min} and $t_{max} = t_{min}$+duration of the data.
- Then in the fitting process, we assume we know t_{min} and t_{max} and offset the initial fitting time to t_{min}
  This is not ideal:
- Unless we can track to real "patient 0" in the population, we can never tell the real $t = 0$, so does t_{min}.
- This also leads to fitting problem that we have seen, if initial β, γ value for fitting is far away from the "real" value, then the mechanism will fail.

As JD mentioned, S_t at any time t should only affect the scaling of the outbreak size since its value only scales the susceptible pool, not the tendency.
Therefore, we can assume $S(t_{min}) = S_{init}$ while t_{min} remains unknown and trying to fit the S_{init} to some level.
Similarly I use notation $I(t_{min}) = I_m$ and $R(t_{min}) = R_m$
We still need I_m and induce $R_m = N - S_m - I_m$ as the initial condition for the fitting simulation.

This leads to one more degree of freedom(now 6, previously 5 as we use $I(0) = I_0$ and naturally assume $R(0) = 0$) in our fitting mechanism.

- β as the transmission rate of underlying SIR model
- γ as the recovery rate of underlying SIR model
- T_B as the baseline testing probability
  - i.e. probability of negative people being tested.
  - Currently assumed to be a constant.
- T_Y as the positive testing probability
  - i.e. probability of positive people being tested.
  - Also a constant
  - Odds ratio between T_B and T_Y is a constant:

$$\Phi = \frac{\frac{T_Y}{1-T_Y}}{\frac{T_B}{1-T_B}}$$

- S_m as the initial condition for S
- I_m as the initial condition for I

In the data we observe and fitting the model to

- Observed testing probability $OT \sim \mathrm{Binom}(N, T)$
- Observed testing positivity $OP \sim \mathrm{Binom}(OT, p)$
- $T(t)$ is the proportion of all tests at time t s.t.

$$T = (1 - Y)T_B + YT_Y$$

  - $Y(t) = I(t)/N$
- $p(t)$ the proportion of positive test at time t s.t.

$$p = \frac{YT_Y}{T} = \frac{1}{1 + (\frac{1}{Y} - 1)\frac{T_B}{T_Y}}$$

  My idea would be deduce an approximation of I_m from initial data and other initial value to "reduce" the number of initial condition to help the fitting machine works better since we need more initial value.

The expectation of OT and OP is $E[OT] = N \times T$ and $E[OP] = OT \times p$ at any time t respectively, and we know OT_m and OP_m from the data. As N and OT should be relatively large, we can use:

$$\hat{T}_m = \frac{OT_m}{N}$$

$$\hat{p} = \frac{OP_m}{OT_m}$$

Then the approximation of I_m could be

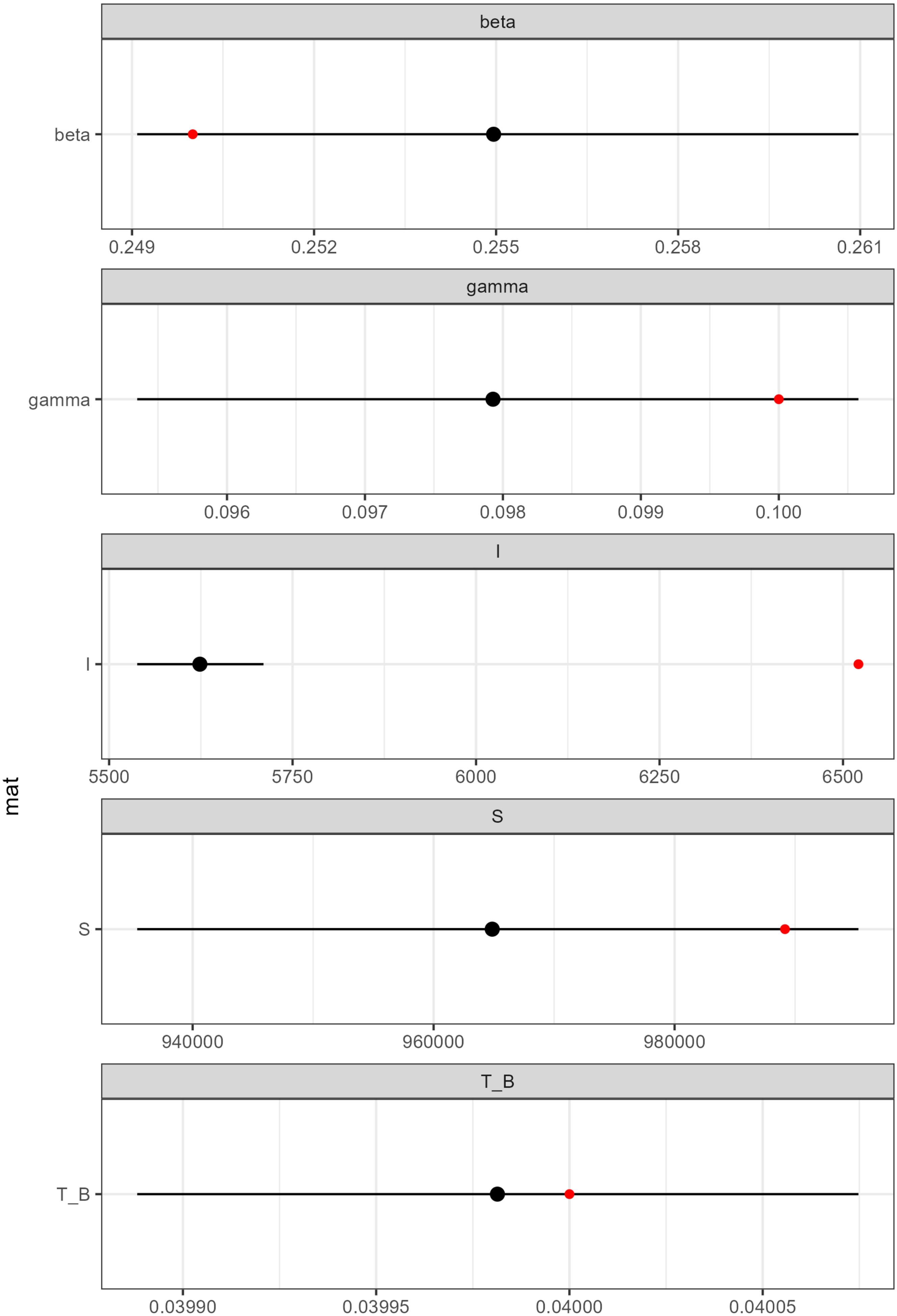
$$\hat{I}_m = \frac{\hat{p}\hat{t}}{T_Y}$$

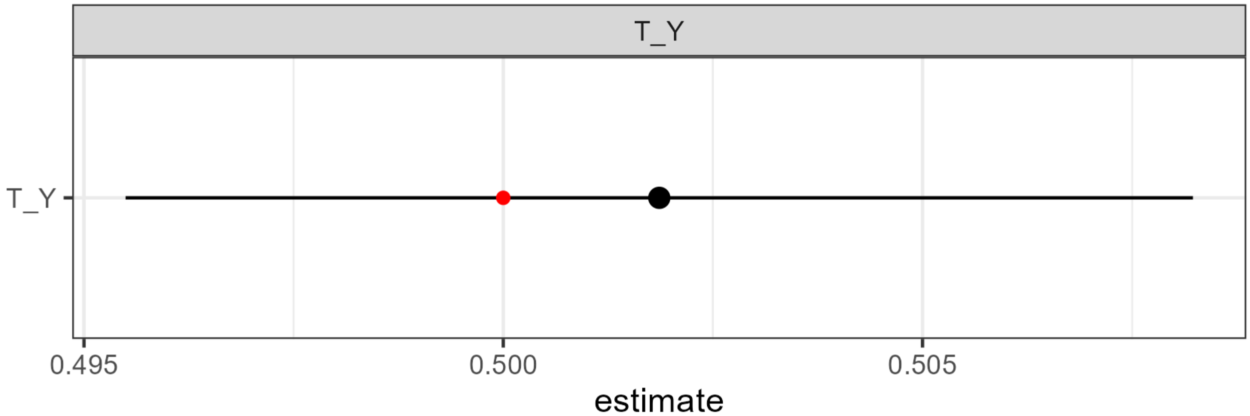
for each T_Y we start with for the fitting.
A constraint would be $S_m + \hat{I}_m \leq N$ but it is not hard to satisfy as S_m is basically free.

A primary implementation is in OR_Sim_SIRwithS_t.R.
The fitting has successfully converge close enough for β, γ, T_Y, T_B with large disturbances to multiple initial value, which previously leads to fitting failure in previous machinery:

- $\beta = \beta_{real} + 0.40$ where $\beta_{real} = 0.25$
- $\gamma = \gamma_{real} + 0.1$ where $\gamma_{real} = 0.1$
- $T_B = 0.04$ is the "real" value
- $T_Y = T_{Yreal} + 0.2$ where $T_{Yreal} = 0.5$
- S_m is 60% of the real value
- $I_m = \hat{I}_m$ is calculated as previously indicated

Fitted parameters comparing with "real" values and 95% CI:


beta
gamma
I
S
T_B
mat
0.249
0.252
0.255
0.258
0.261
0.096
0.097
0.098
0.099
0.100
5500
5750
6000
6250
6500
940000
960000
980000
0.03990
0.03995
0.04000
0.04005



(???) Not very close with I_m, maybe due to randomness of Binomial distribution at low value?

Fitted curves: