Nota técnica - Número de reprodução (R)

SIRD com parâmetros variantes no tempo

- N representa o tamanho da população
- $S(t)$ representa o número de pessoas suscetíveis à doença na população, no dia t. Consideramos $S(t) = N - C(t)$, em que $C(t)$ são os casos confirmados até o dia t
- $D(t)$ representa o número de óbitos por coronavirus até o dia t
- $I(t)$ representa o número de infectados na população no dia t. Consideramos $I(t) = C(t) - D(t) - R(t)$.
- $R(t)$ representa o número de recuperados na população até o dia t. Como não temos, para o Brasil, dados confiáveis de recuperados ao nível de município, usamos a seguinte proxy:

$$R(t) = C(t - 14) - D(t)$$

Utilizamos como base a Alemanha, cujos dados são considerados confiáveis, e verificamos que esse número aproximado de recuperados se aproxima muito do número divulgado.

- β_t é a taxa de transmissão, que deixamos variar no tempo

O sistema fica da seguinte forma:

$$\begin{aligned}
S(t+1) - S(t) &= -\beta_t \cdot \frac{S(t)}{N} \cdot I(t) \\
I(t+1) - I(t) &= \beta_t \cdot \frac{S(t)}{N} \cdot I(t) - \nu_t \cdot I(t) - \mu_t \cdot I(t) \\
R(t+1) - R(t) &= \nu_t \cdot I(t) \\
D(t+1) - D(t) &= \mu_t \cdot I(t)
\end{aligned}$$

Por clareza de interpretação, nossas últimas 2 equações, de forma mais explícita, poderiam ser escritas como:

$$\begin{aligned}
R(t+1) - R(t) &= \gamma_t \cdot \alpha_t \cdot I(t) \\
D(t+1) - D(t) &= \rho_t \cdot (1 - \alpha_t) \cdot I(t)
\end{aligned}$$

em que ρ_t é a taxa de morte (1/tempo médio para a morte), γ_t é a taxa de recuperação e α_t é a probabilidade de recuperação.

Estimação dos parâmetros variantes no tempo

Estamos utilizando um método de estimação por mínimos quadrados suavizados [1].

Reescrevendo a última equação:

$$\Delta D(t+1) = \mu_t \cdot I(t)$$

E temos o seguinte estimador para μ_t:

$$\hat{\mu}_t := \left(\sum_{j=1}^{n} k_{tj} \Delta D(j+1) I(t)\right)\left(\sum_{j=1}^{n} k_{tj} I(j) I(j)\right)^{-1}$$

em que $k_{tj} := K((t-j)/H_\Psi)$, e n é o número de dias (estamos utilizando apenas os dias após o caso 50). Usamos kernel Gaussiano, e $H_\Psi = n^{0.3}$.

De forma análoga para a penúltima equação, temos:

$$\Delta R(t+1) = \nu_t \cdot I(t)$$

$$\hat{\nu}_t := \left(\sum_{j=1}^{n} k_{tj} \Delta R(j+1) I(j)\right)\left(\sum_{j=1}^{n} k_{tj} I(j) I(j)\right)^{-1}$$

E para a primeira equação:

$$\Delta S(t+1) = -\beta_t \frac{S(t)}{N} I(t)$$

$$\hat{\beta}_t := -\left(\sum_{j=1}^{n} k_{tj} \Delta S(j+1) \frac{S(j)}{N} I(j)\right)\left(\sum_{j=1}^{n} k_{tj} \frac{S(j)^2}{N^2} I(j)^2\right)^{-1}$$

Uma observação sobre a sazonalidade dos dados: é sabido que existe uma certa sazonalidade nos números reportados. A média do número de casos é bem menor em feriados e fins de semana, por exemplo. O método de mínimos quadrados suavizados alevia bastante esse problema de sazonalidade. De fato, a nossa estimativa da taxa de transmissão é virtualmente igual para os dados brutos e dados dessazonalizados.

Número básico de reprodução - R_0

Para o número de infectados diminuir ao longo do tempo no modelo SIRD, é necessário que:

$$\beta_t \frac{\cdot S(t)}{N} \cdot I(t) - \nu_t \cdot I(t) - \mu_t \cdot I(t) < 0$$

Logo, se considerarmos o caso com parâmetros fixos no tempo e $t = 0$ (ou seja, $S_t = S_0 \approx N$), temos:

$$\beta < \nu + \mu$$

ou:

$$\frac{\beta}{\nu + \mu} < 1$$

Assim, neste modelo SIRD, o R_0 é dado por $\frac{\beta}{\nu+\mu}$ e um R_0 abaixo de 1 significa que a doença não irá se espalhar ao longo do tempo.

Número efetivo de reprodução - R_t ou R_e

Para avaliar o número efetivo de reprodução ao longo do tempo, consideramos:

- Que a população suscetível diminui ao longo do tempo
- Que os parâmetros β_t, ν_t e μ_t variam ao longo do tempo

$$R_t = \frac{\beta_t}{\nu_t + \mu_t} \cdot \frac{S_t}{N}$$

Referências

[1] Giraitis, Liudas, George Kapetanios, and Tony Yates. "Inference on stochastic time-varying coefficient models." Journal of Econometrics 179.1 (2014): 46-65.