Confidence interval on SEIR predictions

10/04/2024

1 Proof of the method for computing confidence intervals

Assumption :

We suppose that the data of the pandemic observed follows the model h, of parameter $\theta^* \in \mathcal{R}^d$. Let Y_i, $i = 1, \ldots, n$ be the number of hospitalized at each day. We suppose that : $Y_i = h_{\theta^*}(i) + \epsilon_i$, with $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, iid, and independent from all the other variables. The objective is to estimate θ^*. We use $\hat{\theta}$, the least square estimator of θ^* as an estimator of θ^* :

$$\hat{\theta} = \underset{\theta \in \mathbb{R}^d}{\operatorname{argmin}} \sum_{i=1}^{n} (Y_i - h_\theta(i))^2$$

Let :

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$$

$$h_\theta = \begin{pmatrix} h_\theta(1) \\ \vdots \\ h_\theta(n) \end{pmatrix}$$

We have :

$$\hat{\theta} = \underset{\theta \in \mathbb{R}^d}{\operatorname{argmin}} \|Y - h_\theta\|^2$$

Now, if θ is close enough to θ^*, we can write :

$$\forall i \in \{1, ..., n\} : h_\theta(i) = h_{\theta^*}(i) + \nabla_\theta h_{\theta^*}(i)(\theta - \theta^*)$$

which leads to :

$$\hat{\theta} = \underset{\theta \in \mathbb{R}^d}{\operatorname{argmin}} \|Y - h_{\theta^*}(i) - \nabla_\theta h_{\theta^*}(i)(\theta - \theta^*)\|^2$$

Let us define :

$$\tilde{Y} = Y - h_{\theta^*}$$
$$\beta = \theta - \theta^*$$
$$\hat{\beta} = \theta - \hat{\theta}$$

and let us define the matrix $A \in \mathbb{R}^{n \times d}$ such that $\forall i \in \{1, ..., n\}, \forall j \in \{1, ..., d\}, A_{i,j} = \frac{dh_{\theta^*}}{d\theta_j}(i)$.
The previous problem can be re-written as :

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^d}{\operatorname{argmin}} \left\|\tilde{Y} - A\beta\right\|^2$$

This is a regression linear problem.

Let us solve this problem in the general case. Let $\tilde{Y}_i = A_i\beta^* + \epsilon'_i$, with $\epsilon'_i \sim \mathcal{N}(0, \sigma'^2)$. The solution of this problem is explicitly :

$$\hat{\beta} = (A^T A)^{-1} A^T \tilde{Y}$$

This least-square estimator is unbiased :

$$\mathbb{E}[\hat{\beta}] = \beta^*$$

$$\hat{\beta} = \left(\frac{1}{n}\sum_{i=1}^{n} A_i^T A_i\right)^{-1} \times \left(\frac{1}{n}\sum_{i=1}^{n} A_i^T \tilde{Y}_i\right)$$

Let us note :

$$\hat{D} = \frac{1}{n}\sum_{i=1}^{n} A_i^T A_i, \quad \text{and} \quad \hat{\delta} = \left(\frac{1}{n}\sum_{i=1}^{n} A_i^T \tilde{Y}_i\right)$$

We have :

$$\hat{\beta} = \hat{D}^{-1}\hat{\delta}$$

$$\hat{D} \underset{a.s}{\to} \mathbb{E}[A_i^T A_i]$$

$$\hat{\delta} \underset{a.s}{\to} \mathbb{E}[A_i^T \tilde{Y}_i]$$

$$\hat{\beta} = \hat{D}^{-1}\hat{\delta} \underset{a.s}{\to} D^{-1}\delta$$

as $\phi : A \to A^{-1}$ is continuous on $\mathcal{GL}_n(\mathbb{R})$.
Now, let us show that $\hat{\beta}$ is asymptotically normal :

$$\begin{aligned}
\sqrt{n}(\hat{\beta} - \beta^*) &= \sqrt{n}(\hat{D}^{-1}\hat{\delta} - \beta^*) \\
&= \sqrt{n}(\hat{D}^{-1}\hat{\delta} - \hat{D}^{-1}\hat{D}\beta^*) \\
&= \sqrt{n}\hat{D}^{-1}(\hat{\delta} - \hat{D}\beta^*) \\
&= \sqrt{n}\hat{D}^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} A_i^T \tilde{Y}_i - \frac{1}{n}\sum_{i=1}^{n} A_i^T A_i \beta^*\right) \\
&= \frac{\sqrt{n}}{n}\hat{D}^{-1}\left(\sum_{i=1}^{n} A_i^T (\tilde{Y}_i - A_i\beta^*)\right) \\
&= \frac{1}{\sqrt{n}}\hat{D}^{-1}\left(\sum_{i=1}^{n} A_i^T \epsilon_i\right)
\end{aligned}$$

This line is made of two terms. Let's show that each one of them converges in law.

$$\begin{aligned}
\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n} A_i^T \epsilon'_i\right) &= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} A_i^T \epsilon'_i\right) \\
&= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} A_i^T \epsilon'_i - 0\right) \\
&\xrightarrow{a.s.} \mathcal{N}(0, \mathrm{Var}(A_i^T \epsilon_i))
\end{aligned}$$

Yet, as $\epsilon_i \perp A_i$ and $\mathbb{E}[A_i^T \epsilon_i'] = 0$, $\text{Var}(A_i^T \epsilon_i) = \mathbb{E}[A_i A_i^T \epsilon_i^2] = \mathbb{E}[A_i A_i^T]\sigma'^2$.
Finally, $\frac{1}{\sqrt{n}} \left(\sum_{i=1}^n A_i^T \epsilon_i'\right) \xrightarrow{a.s.} \mathcal{N}(0, D\sigma'^2)$.
On the other hand, $\hat{D}^{-1} \xrightarrow{\mathcal{L}} D^{-1} = Cte$.
Finally, with Slutsky, we obtain that :

$$\begin{aligned}\sqrt{n}(\hat{\beta} - \beta^*) &\xrightarrow{\mathcal{L}} D^{-1}\mathcal{N}(0, D\sigma'^2) \\ &\xrightarrow{\mathcal{L}} \mathcal{N}(0, D^{-1}\sigma'^2) \\ &\xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma'^2(A^T A)^{-1})\end{aligned}$$

Let's get back to the first problem :
As $\beta^* = 0$ and $\hat{\beta} = \hat{\theta} - \theta^*$, we have :

$$\sqrt{n}(\hat{\theta} - \theta^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2(A^T A)^{-1})$$

and,

$$\hat{\theta} \sim \mathcal{N}(\theta^*, \frac{\sigma^2}{n}(A^T A)^{-1})$$

As a first conclusion, we have that $\hat{\theta}$ is asymptotically normal.
Let Σ be the covariance matrix estimated from the computation of $\hat{\theta}$. In our case, $\Sigma = \frac{\sigma^2}{n}(A^T A)^{-1}$.
As $\hat{\theta}$ is asymptotically normal, we can apply the delta-method :

$$\sqrt{n}(\hat{\theta} - \theta^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\ \Sigma)$$

$$\sqrt{n}(h_{\hat{\theta}} - h_{\theta^*}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nabla_\theta h_\theta^T \Sigma \nabla_\theta h_\theta)$$

And finally :

$$h_{\hat{\theta}} \rightarrow \mathcal{N}(h_{\theta^*}, \frac{1}{n}\nabla_\theta h_\theta^T \Sigma \nabla_\theta h_\theta)$$

By estimating $\frac{1}{n}\Sigma$ from `curve_fit`, we can compute the confidence interval of the prediction with the quantiles of the normal distribution.