Confidence interval on SEIR predictions

10/04/2024

We observe data $(Y_i)_{i\in[\![1,n]\!]}$, which corresponds to the number of deaths per day.

Assumption :

The number of deaths per day follows a SEIR model :
$Y_i = h_\theta(i) + \epsilon_i$

With $\theta = (\beta, \gamma, d) \in \mathbb{R}^3$ the parameters of the SEIR model and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, i.i.d with σ unknown.

We want to find θ and to compute confidence intervals on the predictions of the SEIR model.

Let $Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$

We find $\hat{\theta}$, an estimator of θ by minimizing the difference between h_θ and the observed data.

$\hat{\theta} = \underset{\tilde{\theta}\in\mathbb{R}^3}{\operatorname{argmin}} \left(\sum_{i=1}^{n} (h_{\tilde{\theta}}(i) - Y_i)^2\right)$
As θ is fixed with $(Y_i)_{i\in[\![1,n]\!]}$ fixed, we note $\hat{\theta} = f(Z)$

Thanks to the Δ-method, we have :

$Var(\hat{\theta}) = \nabla_\theta^T f(Z) Cov(Z) \nabla_\theta f(Z)^T$, with $\nabla_\theta f(Z)$ the hessian of f evaluated in Z.

We also have, as the ϵ_i are independante gaussian variables :
$Cov(Z) = \sigma^2 \mathbf{I}_n$

We estimate σ^2 like this :

$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} \epsilon_i^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - h_{\hat{\theta}}(i))^2$

We can now have access to $Var(\hat{\theta})$ and approximate it with a normal distribution : $\mathcal{N}(\hat{\theta}, Var(\hat{\theta}))$

We can now sample different θ_i from this distribution and predict the behaviour of the pandemic for each of this value of θ. It gives us a confidence interval on the predictions of the SEIR model.

We can also compute the confidence interval directly on a prediction with the formula :

$Var(h_{\hat{\theta}}(i)) = \nabla_\theta^T h_{\hat{\theta}}(i) Var(\hat{\theta}) \nabla_\theta h_{\hat{\theta}}(i)$