Modelling Epidemic Trends

Statistical problem

Suppose we observe a sample of data $\mathcal{N}_i = \{N_{it} : 1 \leq 1 \leq T\}$ for n different countries i, where N_{it} is the number of new infections on day t in country i.[1] We assume that $N_{it} \sim \mathcal{P}_{\lambda_i(t/T)}$, that is, N_{it} is Poisson distributed with (time-varying) intensity parameter $\lambda_i(t/T)$. For simplicity, we suppose that the random variables N_{it} are independent both across countries i and time t. In the current Covid-19 crisis, independence across countries i seems justified as borders between countries are effectively closed. Independence across t is more debatable, but may be justified as follows: Since $\lambda_i(t/T) = \mathbb{E}[N_{it}]$, we obtain the nonparametric regression model

$$N_{it} = \lambda_i\Big(\frac{t}{T}\Big) + u_{it}, \qquad (1)$$

where $u_{it} = N_{it} - \mathbb{E}[N_{it}]$ with $\mathbb{E}[u_{it}] = 0$. The persistence in the time series process $\mathcal{N}_i$ is generated by the time-varying trend $\lambda_i(t/T)$ in model (1). In the simplest version of the model, one may assume the noise terms u_{it} to be independent across t, thus neglecting potential dependencies that may cause further persistence.

Our aim is to test whether the dynamics of the Poisson process $\mathcal{N}_i = \{N_{it} : 1 \leq 1 \leq T\}$ are the same across countries i, that is, whether the time-varying intensity function λ_i has the same shape across i. Suppose we are particularly interested in comparing a specific country i with all other countries j in our sample. In this case, the null hypothesis to be tested can be formulated as

$$H_0^{(i)} : \lambda_i(\cdot) = \lambda_j(\cdot) \text{ for all } j \in \{1, \dots, n\} \text{ with } j \neq i.$$

If we are interested in comparing all countries with each other, the null hypothesis becomes

$$H_0 : \lambda_i(\cdot) = \lambda_j(\cdot) \text{ for all } i, j \in \{1, \dots, n\} \text{ with } j \neq i.$$

In what follows, we focus on the null hypothesis $H_0^{(i)}$ because it requires a bit less notation, but our approach immediately carries over to the hypothesis H_0.

[1] We can also work with other kinds of count data, e.g., with the accumulated number of infections and the number of deaths (per day or accumulated).

Testing $H_0^{(i)}$

Construction of the test statistic:

(1) Let $\{\mathcal{I}_k : 1 \le k \le K\}$ be a family of subintervals of $[0,1]$. We first define a statistic to test the hypothesis $H_0^{(ijk)}$ that λ_i is equal to λ_j on the subinterval $\mathcal{I}_k$, that is,

$$H_0^{(ijk)} : \lambda_i(u) = \lambda_j(u) \text{ for all } u \in \mathcal{I}_k.$$

To do so, we first introduce the expression

$$\hat{s}_{ijk,T} = \sum_{t=1}^{T} w_k\Big(\frac{t}{T}\Big)(N_{it} - N_{jt}),$$

where $w_k(t/T)$ is a (rectangular) kernel weight defined by

$$w_k\Big(\frac{t}{T}\Big) = \frac{\mathbf{1}(\frac{t}{T} \in \mathcal{I}_k)}{\{\sum_{s=1}^{T} \mathbf{1}(\frac{s}{T} \in \mathcal{I}_k)\}^{1/2}}.$$

Since N_{it} is Poisson distributed with intensity $\lambda_i(t/T)$, it holds that $\mathbb{E}[N_{it}] = \mathrm{Var}(N_{it}) = \lambda_i(t/T)$ and thus $\mathrm{Var}(u_{it}) = \lambda_i(t/T)$. This implies the following: If $\lambda_i = \lambda_j$ on $\mathcal{I}_k$, then

$$\begin{aligned}
\sigma^2_{ijk,T} := \mathrm{Var}(\hat{s}_{ijk,T}) &= \sum_{t=1}^{T} w_k^2\Big(\frac{t}{T}\Big)\mathbb{E}\big[(u_{it} - u_{jt})^2\big] \\
&= \sum_{t=1}^{T} w_k^2\Big(\frac{t}{T}\Big)\Big\{\lambda_i\Big(\frac{t}{T}\Big) + \lambda_j\Big(\frac{t}{T}\Big)\Big\} \\
&\approx h_k^{-1} \int_{u \in \mathcal{I}_k} \{\lambda_i(u) + \lambda_j(u)\} du,
\end{aligned}$$

where h_k is the length of the interval $\mathcal{I}_k$. In order to normalize the variance of the statistic $\hat{s}_{ijk,T}$, we scale it by an estimator of $\sigma_{ijk,T}$. In particular, we use the expression $\hat{\psi}_{ijk,T} := \hat{s}_{ijk,T}/\hat{\sigma}_{ijk,T}$ as a test statistic, where

$$\hat{\sigma}^2_{ijk,T} = \sum_{t=1}^{T} w_k^2\Big(\frac{t}{T}\Big)\{N_{it} + N_{jt}\}.$$

The overall test statistic thus has the form

$$\hat{\psi}_{ijk,T} := \frac{\hat{s}_{ijk,T}}{\hat{\sigma}_{ijk,T}} = \frac{\sum_{t=1}^{T} \mathbf{1}(\frac{t}{T} \in \mathcal{I}_k)(N_{it} - N_{jt})}{\{\sum_{t=1}^{T} \mathbf{1}(\frac{t}{T} \in \mathcal{I}_k)(N_{it} + N_{jt})\}^{1/2}}.$$

(2) We next combine the test statistics $\hat{\psi}_{ijk,T}$ for all intervals $\{\mathcal{I}_k : 1 \le k \le K\}$ and all countries j to obtain the following multiscale test statistic of the hypothesis $H_0^{(i)}$:

$$\hat{\Psi}_{n,T}^{(i)} = \max_{j \neq i} \max_{1 \le k \le K} \Big\{ \big| \hat{\psi}_{ijk,T} \big| - p(h_k) \Big\},$$

where as above h_k is the length of the interval $\mathcal{I}_k$ and $p(h) = \sqrt{2 \log(1/h)}$.

Computation of the critical value:

(1) Under $H_0^{(i)}$ and appropriate regularity conditions, we obtain that

$$\begin{aligned} \hat{\psi}_{ijk,T} &= \frac{\sum_{t=1}^T \mathbf{1}(\frac{t}{T} \in \mathcal{I}_k)(u_{it} - u_{jt})}{\{\sum_{t=1}^T \mathbf{1}(\frac{t}{T} \in \mathcal{I}_k)(N_{it} + N_{jt})\}^{1/2}} \\ &\approx \frac{1}{\sqrt{2T}} \sum_{t=1}^T \mathbf{1}\Big(\frac{t}{T} \in \mathcal{I}_k\Big) \Big\{ \frac{u_{it}}{\sqrt{\lambda_i(\frac{t}{T})}} - \frac{u_{jt}}{\sqrt{\lambda_j(\frac{t}{T})}} \Big\}, \end{aligned}$$

where the variables $u_{it}/\sqrt{\lambda_i(t/T)}$ and $u_{jt}/\sqrt{\lambda_j(t/T)}$ have zero mean and unit variance.

(2) We now define a Gaussian version of the statistic displayed in the last line above. In particular, we let

$$\hat{\phi}_{ijk,T} = \frac{1}{\sqrt{2T}} \sum_{t=1}^T \mathbf{1}\Big(\frac{t}{T} \in \mathcal{I}_k\Big) \big\{ Z_{it} - Z_{jt} \big\},$$

where Z_{it} are i.i.d. standard normal random variables for $1 \le t \le T$ and $1 \le i \le n$. With this, we define the statistic

$$\Phi_{n,T}^{(i)} = \max_{j \neq i} \max_{1 \le k \le K} \Big\{ \big| \phi_{ijk,T} \big| - p(\ell_k) \Big\}$$

and denote its $(1-\alpha)$-quantile by $q_{n,T}^{(i)}(\alpha)$.

Our multiscale test of the hypothesis $H_0^{(i)}$ is now carried out as follows: For a given significance level $\alpha \in (0,1)$, reject $H_0^{(i)}$ if $\hat{\Psi}_{n,T}^{(i)} > q_{n,T}^{(i)}(\alpha)$.