Nowcasting Covid-19 onset in the UK

Ryan Teo

2022-06-16

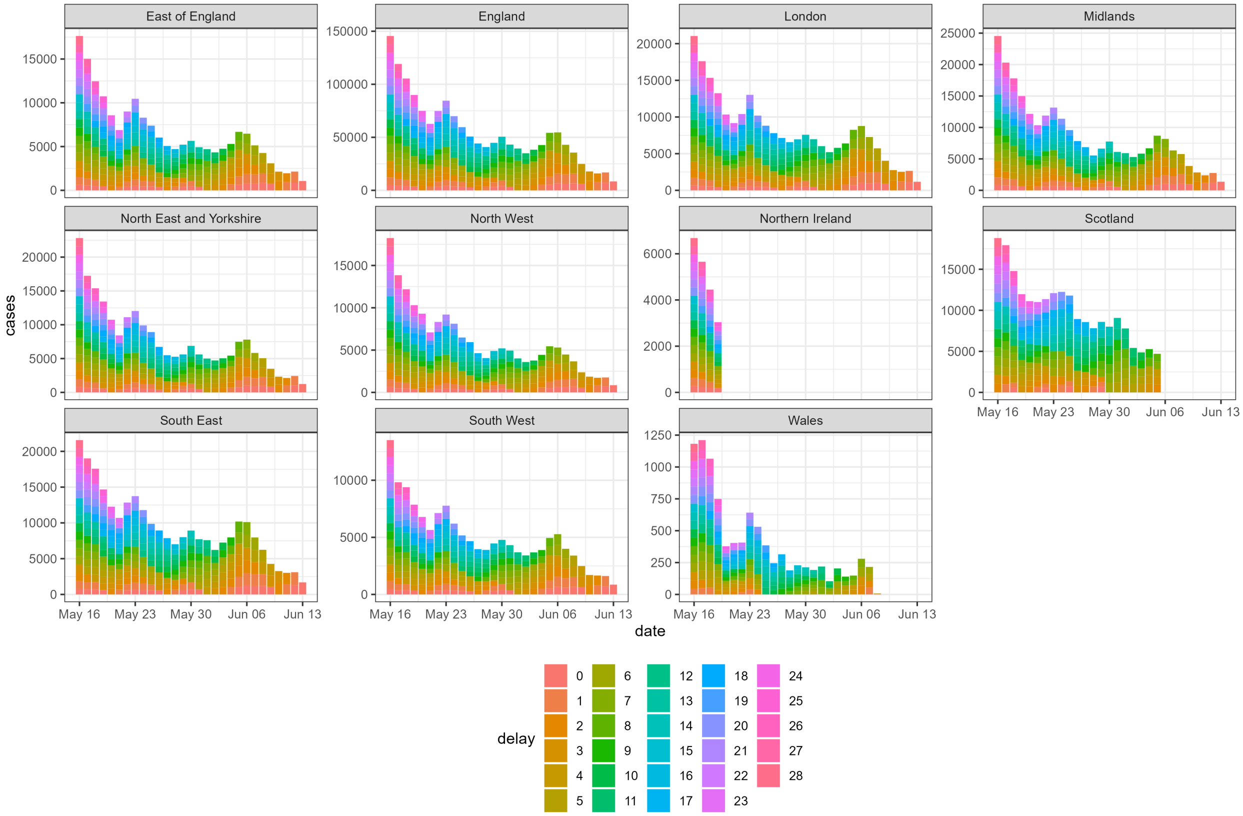
15/06/2022

Software

- Successfully installed `epinowcast`, managed to run the nowcast for hospitalisations in Germany
- Set up C++ tool chain for `CmdStan` v2.29.2

Data

- Downloaded `newCasesBySpecimenDate` from the UK Covid-19 dashboard
  - Aggregated by LTLAs then NHS England regions
  - Age stratification also available `newCasesBySpecimenDateAgeDemographics`
  - What is a suitable time frame?
- Plotted cases over time by reporting delay
  - Scotland, Wales, and Northern Ireland seem to be working on different reporting schedules
  - What other visualisations are helpful?
  - Would be interesting to see if the distribution of delays changed over time maybe due to testing advice



Theory

The nowcasting problem is described as follows (Höhle and Heiden 2014).

- $n_{t,d}$: number of cases with
  - onset on day $t, t = 0, ..., T$
  - reported with a delay of $d, d = 0, ..., D$ days, i.e. reports arrive on day $t + d$.
    - We may assume that delays can occur only up to a maximum of D days
- We observe $\ltimes = (n_{t,d} : (t, d) \in A^m_T)$ where

$$\{(t, d) : \max(T - m, 0) \leq t \leq T, 0 \leq d \leq \min(D, T - t)\}$$

  is the right-angle trapezoidal observation region at time T when using a moving window of size m.
- $N(t, T) = \sum_{d=0}^{\min(T-t,D)} n_{t,d}$: number of cases which occured on t which are reported until time T
  - The aim of nowcasting is to predict the total number of cases which occurred on $t, t = T - D, ..., T$ given the information at time T
- Reporting delay (in days) of a case occurring at time t follows a distribution with probability mass function $f_t(d) = p_{t,d}, d = 0, ..., D$, where $\sum_{d=0}^{D} p_{t,d} = 1$.
  - Considering time t and conditioning on number of cases $N(t, T)$ observed by time T, the observed $n'_{t,d}s$ in the reporting triangle due to the right-truncation have a multinomial distribution with size $N(t, \infty)$ and cell probabilities $p'_{t,d} = p_{t,d} / \sum_{i=0}^{\min(D,T-t)} p_{t,i}$ for $d = 0, ..., \min(D, T - t)$.
- If we assume that occurrence of new cases follow an underlying inhomogeneous Poisson process, $N(t, \infty) \sim \text{Poisson}(\lambda_t), t = \max(0, T - m), ..., T$, the above conditional formulation of the $n'_{t,d}s$ originating from multinomial sampling can be re-formulated as an unconditional likelihood corresponding to an incomeplete contingency table

$$n_{t,d} \sim \text{Poi}(\mu_{t,d}), \quad (t, d) \in A^m_T, \text{ where } \log(\mu_{t,d}) = \log(\lambda_t) + \log(p_{t,d})$$

- Nowcasting can thus be divided into

1. Determining the $\lambda_t' s$
2. Determining the sequence of $p_{t,d}' s$
3. Predicting the unobserved $n_{t,d}' s$ to compute the total $N(t, \infty)$

Höhle, Michael, and Matthias an der Heiden. 2014. "Bayesian Nowcasting During the STEC O104:H4 Outbreak in Germany, 2011." *Biometrics* 70 (4): 993–1002. https://doi.org/10.1111/biom.12194.