One-pager on R_t estimation with geographical dropout

KP, DJM

Last revised: October 29, 2024

1 As presented at the conference

- Let $y_{\ell,t}$ be incidence in location ℓ at time t
- Let y_t be observed incidence over the region with $y_t = \sum_\ell y_{\ell,t}$
- We have regional estimates of R_t from time $t = 1, \ldots, T + H$.
- There is a subregion ℓ with data available only up until time t.
- Define $\Lambda_t = \sum_{i=1}^t \omega_i y_{t-i}$ to be the convolved incidence in the region and $\Lambda_{\ell,t}$ be the same for location ℓ

The "deterministic dropout correction" is:

Input $\{R_t\}_{t=1}^{T+H}$, $\{R_{\ell,t}\}_{t=1}^{T}$, $\Lambda_{\ell,t}$.
For $h = 1, \ldots, H$, do

1. Predict $\hat{y}_{\ell,T+h} = R_{T+h}\Lambda_{\ell,T+h-1}$ (set to it's expected value)
2. Convolve $\Lambda_{\ell,T+h} = \sum_{i=1}^{T+h} \omega_i \tilde{y}_{T+h-i}$, where $\tilde{y}_j = y_j \mathbb{1}[j \leq T] + \hat{y}_j \mathbb{1}[j > T]$.
3. Estimate local $\hat{R}_{\ell,T+h} = R_{T+h}\frac{\Lambda_{\ell,T+h-1}}{\Lambda_{\ell,T+h}}$.

2 As actually implemented

Once we started implementing, a few minor changes were necessary.

- We don't actually use $\hat{R}_{\ell,T+h}$ for anything. So, we don't calculate it.
- Because location ℓ is unavailable for $t = T + 1, \ldots, T + H$, we also don't have Λ_t for $t = T + 1, \ldots, T + H$. For $h \geq 1$ we used the correction $1 - \Lambda_{\ell,t}/\Lambda_t$, to rescale regional incidence (and convolved incidence).
- Once we have the sequence $\{\hat{y}_{\ell,t}\}$ for $t = 1, \ldots, T + H$, we just throw that into the local R_t estimation routine.

3 Suggested implementation

We are predicting $\hat{y}_{\ell,T+h}$ using R_{T+h} and $\Lambda_{\ell,T+h-1}$, both of which are smooth. Then we reconvolve and iterate. It may be more productive to view this as a process with 2 modules, a `forecaster` and an `Rtestimator`. Then the meta procedure is

1. Use `forecaster` to produce $\{\Lambda_{\ell,T+h}\}$ for $1 \leq h \leq H$. We do this conditional on all available information (incidence at the current location, the region, or whatever we want in between). We can even use auxiliary signals (wastewater or similar). Our current version is just a very simple `forecaster` (uses global R_t and local convolved incidence).
2. Calculate $\hat{y}$ by taking the first differences of $\{\Lambda_{\ell,T+h}\}$.
3. Use `Rtestimator` to produce $\{R_{\ell,t}\}_{t=1}^{T+H}$