GeiGovid

April 7, 2020

Abstract

Airborne infectious diseases are an invisible threat. This paper sketches an infectious diseases geiger counter that 'runs ahead' of the disease, revealing potentially dangerous exposure to infectious diseases while respecting privacy.

1 Overview

The idea is to transmit a probabilistic measure of exposure by Bluetooth beacons, upon which sensible estimates can be given for how likely a person is to be infected.

Exposing such a tangible, roughly accurate measure for the exposure may increase awareness and adherence to social distancing.

The next section illustrates a rough derivation of what such a probability estimate could look like:

Simplifying assumptions are made: The infection is 'memoryless' in that after a person is infected, exposure to other infected hosts does not change the prognosis of the disease.

2 Infection model

In the following we first build a model in discrete time of infection probability from the perspective of a single person P, assuming our infection can be modelled as independent Bernoulli trials that are not necessarily identically distributed. In order to do this we pick a timeframe of interest that spans from some time in the past $t - T$ to the current time t, outside of which we assume no infection could occur. This timeframe is subdivided into n intervals:

$$k \in \{1, ..., n+1\} : t_k = t - T + \frac{k-1}{n-1}T$$
$$\delta t = t_{k+1} - t_k$$

Our probability space is thus

$$\mathcal{X} = \{0,1\}^n$$
$$\mathcal{A} = 2^{\mathcal{X}}$$
$$p_k \in [0,1] : \ \mathbb{P}((x_1,...,x_n)) = \prod_{k=1}^{n}(x_k p_k + (1-x_k)(1-p_k))$$

where $x_k = 1$ means an infection occurred in the interval t_k and p_k is the probability of infection in that interval.

This probability itself can be expressed as the probability of any nearby people infecting our person of interest, which is another sequence of Bernoulli trials which are independent, not necessarily identically distributed. l indexes all other people, in practice of course most of which will not be able to infect P due to distance:

$$\begin{aligned} p_k &= \mathbb{P}'\left(\bigcup_{l=1}^{N} \overbrace{A_{kl}}^{l \text{ infects } P \text{ at interval } k}\right) \\ &= 1 - \mathbb{P}'\left(\bigcap_{l=1}^{N} A_{kl}^c\right) \\ &= 1 - \prod_{l=1}^{N} \mathbb{P}'\left(A_{kl}^c\right) \\ &= 1 - \prod_{l=1}^{N} (1 - p_{kl}) \end{aligned}$$

Assuming nice enough functions with sufficient differentiability

$$p_{kl} = p_l[k] = \mathcal{F}_l\left(t_{k+1}\right) - \mathcal{F}_l\left(t_k\right)$$

Note $\mathcal{F}$ isn't a cumulative distribution function, as it doesn't always hold that $\mathcal{F}(t) = 1$, however monotonicity and positivity do hold.

$$\lim_{\delta t \to 0} \frac{\mathcal{F}_l\left(t_k + \delta t\right) - \mathcal{F}_l\left(t_k\right)}{\delta t} = \mathcal{P}_l\left(t_k\right)$$

Thus for sufficiently small δt it holds that

$$p_{kl} \approx \mathcal{P}_l(t_k)\delta t$$

3 Probability of infection

$$\begin{aligned}
\mathbb{P}(\text{not infected}) &= \prod_{j=1}^{n}(1-p_j) && (1)\\
&= \prod_{j=1}^{n}\prod_{k=1}^{N}(1-p_{jk}) && (2)\\
&= \prod_{k=1}^{N}\prod_{j=1}^{n}(1-p_{jk}) && (3)\\
&= \prod_{k=1}^{N}\prod_{j=1}^{n}\left(1-\mathcal{P}_k\left(t_j\right)\delta t+\mathcal{O}\left(\delta t^2\right)\right) && (4)\\
&= \prod_{k=1}^{N}\exp\left(\sum_{j=1}^{n}\ln\left(1-\mathcal{P}_k\left(t_j\right)\delta t+\mathcal{O}\left(\delta t^2\right)\right)\right) && (5)
\end{aligned}$$

Since

$$|x| \leq D < 1 : \ln(1-x) = -\sum_{l=1}^{\infty}\frac{x^l}{l}$$

for sufficiently small δt we have absolute convergence

$$\begin{aligned}
\sum_{j=1}^{n}\ln\left(1-\mathcal{P}_k\left(t_j\right)\delta t+\mathcal{O}\left(\delta t^2\right)\right) &= -\sum_{j=1}^{n}\sum_{l=1}^{\infty}\frac{\left(\mathcal{P}_k\left(t_j\right)\delta t+\mathcal{O}\left(\delta t^2\right)\right)^l}{l}\\
&= -\sum_{j=1}^{n}\left(\mathcal{P}_k\left(t_j\right)\delta t+\mathcal{O}\left(\delta t^2\right)\right)\\
&\overset{\delta t\to 0}{=} -\int_{t-T}^{t}\mathcal{P}_k(\tau)\,d\tau
\end{aligned}$$

Inserting back into equation 5 gives

$$\mathbb{P}(\text{not infected}) = \exp\left(-\int_{t-T}^{t}\sum_{k=1}^{N}\mathcal{P}_k(\tau)\,d\tau\right)$$

4 Connecting infectiousness with probability of infection

We'll use the expectation of an infectiousness random variable as our estimate of current infectiousness, again first in discrete time:

$$\begin{aligned}
\mathfrak{i} &= E\left[I[n]\right]\\
&= E\left[E\left[I[n]|S=j\right]\right]
\end{aligned}$$

Here S denotes the time of infection.
Expand into

$$\sum_{j=1}^{n} E\left[I[n]|S=j\right] \cdot \mathbb{P}(S=j)$$

Now we use the assumption that the prognosis of disease is independent of subsequent infections, so only the first counts, from then on which we know from statistical data how the expectation of I develops:

$$\sum_{j=1}^{n} E\left[I[n]|S=j\right] \cdot \mathbb{P}(S=j) = \sum_{j=1}^{n} \mathcal{I}\left(t_n - t_j\right) \cdot \mathbb{P}(S=j) \tag{6}$$

The probability that the infection started at j means all prior Bernoulli trials evaluated to 0:

$$\begin{aligned}
\mathbb{P}(S=j) &= p_j \prod_{k=1}^{j-1}(1-p_k) \\
&= \left(\overbrace{\sum_{l=1}^{N} \mathcal{P}_l\left(t_j\right) \delta t + \mathcal{O}\left(\delta t^2\right)}^{\theta(t_j)} \right) \prod_{k=1}^{j-1} \left(1 - \sum_{l=1}^{N} \mathcal{P}_l\left(t_k\right) \delta t + \mathcal{O}\left(\delta t^2\right) \right) \\
&= \left(\theta\left(t_j\right) \delta t + \mathcal{O}\left(\delta t^2\right)\right) \exp\left(\sum_{k=1}^{j-1} \ln\left(1 - \theta\left(t_k\right) \delta t + \mathcal{O}\left(\delta t^2\right)\right) \right) \\
&= \left(\theta\left(t_j\right) \delta t + \mathcal{O}\left(\delta t^2\right)\right) \exp\left(- \sum_{k=1}^{j-1} \left(\theta\left(t_k\right) \delta t + \mathcal{O}\left(\delta t^2\right)\right) \right) \\
&= \left(\theta\left(t_j\right) \delta t + \mathcal{O}\left(\delta t^2\right)\right) \exp\left(- \sum_{k=1}^{j-1} \theta\left(t_k\right) \delta t + \mathcal{O}\left(\delta t\right) \right) \\
&= \left(\theta\left(t_j\right) \delta t + \mathcal{O}\left(\delta t^2\right)\right) \left(\exp\left(- \sum_{k=1}^{j-1} \theta\left(t_k\right) \delta t \right) + \mathcal{O}\left(\delta t\right) \right) \\
&= \theta\left(t_j\right) \exp\left(- \sum_{k=1}^{j-1} \theta\left(t_k\right) \delta t \right) \delta t + \mathcal{O}\left(\delta t^2\right)
\end{aligned}$$

Inserting into equation 6 gives

$$\sum_{j=1}^{n} \mathcal{I}\left(t_n - t_j\right) \theta\left(t_j\right) \exp\left(- \sum_{k=1}^{j-1} \theta\left(t_k\right) \delta t \right) \delta t + \mathcal{O}\left(\delta t\right)$$

$$\int_{t-T}^{t} \mathcal{I}(t-\tau) \theta(\tau) \exp\left(- \int_{t-T}^{\tau} \theta\left(\tau'\right) d\tau' \right) d\tau$$

5 Privacy issues

None beyond any introduced by Bluetooth.

6 Robustness against DoS

DoS can be attempted in three ways: An attacker can choose to broadcast a different estimator, either higher or lower than the protocol would suggest.

An estimate that is broadcasted cannot be exceedingly high as this risks being detected as out of the ordinary. The more serious attack is thus to send inconspicuous estimates, yet fake a large number of devices. This seems not entirely riskless, but plausible.

The other option is to send an estimate that is too low. No privacy-preserving system that protects participants from being identified against their will can prevent this. The worst thing an attacker can do in this case is to not participate at all. This is acceptable. Unacceptable would be an attacker able to disrupt the operation of the system for honest users.

7 Benefits

- Easy to implement
- Low footprint
- Immunity can be advertised with a zero estimator

8 Downsides

- Technology-tyranny, opaque to end user
  - Correctness of contact tracing easy to demonstrate
- Relies on statistical information that may be unavailable
- Needs external jump start
- Might need manual intervention if model gets out of hand

9 Extensions

An extension could try closing the feedback loop to the health authority in a privacy-respecting way, to convey information which can be used to refine the model. One half of the loop is to retrieve updated models from the health authority and updating prior information. Adding model version to the beacon could add accuracy, but complicates reasoning about attack vectors. Less accurate, but more robust would be to assume information collected is using the most up to date model available from the health authority.

10 Caveats

Not reviewed by anybody with experience in infectious diseases. Or any other expert.

Appendix