Normal Approximations for Occupancy Processes using Stein's method

Liam Hodgkinson liam.hodgkinson@uqconnect.edu.au
Ross McVinish r.mcvinish@uq.edu.au
Philip K. Pollett pkp@maths.uq.edu.au

Occupancy Processes

A discrete-time Markov chain $\boldsymbol{X}_t = (X_{i,t})_{i=1}^n$ on $\{0,1\}^n$

- number of nodes n
- given $\boldsymbol{X}_t = \boldsymbol{x}$, $X_{1,t+1}, \dots, X_{n,t+1}$ are independent
- potentially time-inhomogeneous
- dynamics dictated by a one-step global rule $\boldsymbol{P}_t = (P_{i,t})_{i=1}^n$:

$$P_{i,t}(\boldsymbol{x}) := \mathbb{P}(X_{i,t+1} = 1 \mid \boldsymbol{X}_t = \boldsymbol{x})$$

- general class of processes with examples appearing in **ecology**, epidemiology (see below), **physics, computer science, social science**...

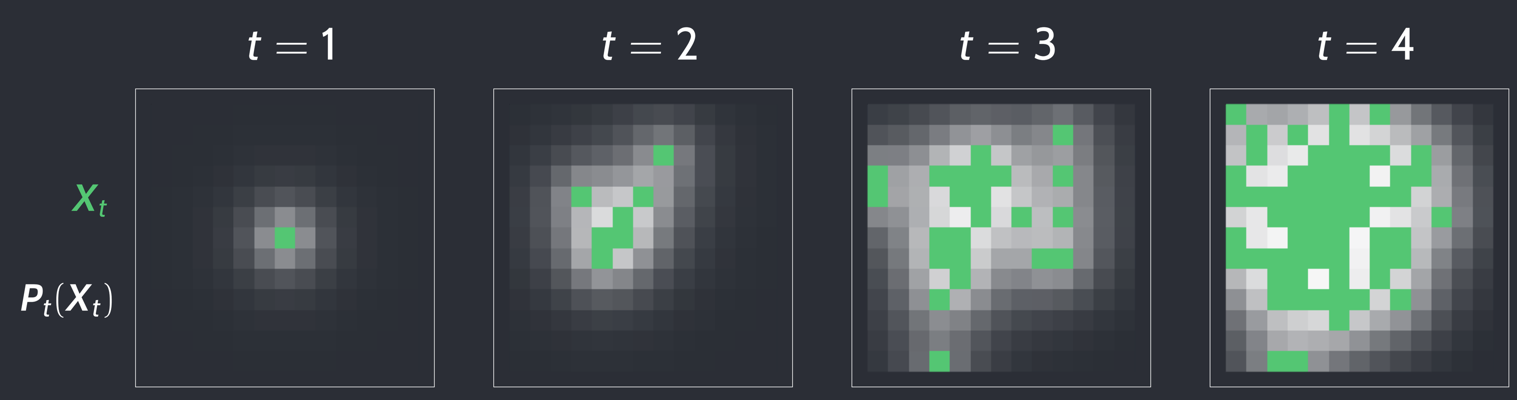

Figure 1: *An example of an occupancy process on a 13 × 13 grid. The process X_t is represented in green with the underlying heat map representing the probabilities $P_{i,t}(X_t)$ of each node of the process for the next time point. Here, i indexes grid points.*

A European Epidemic (SIS)

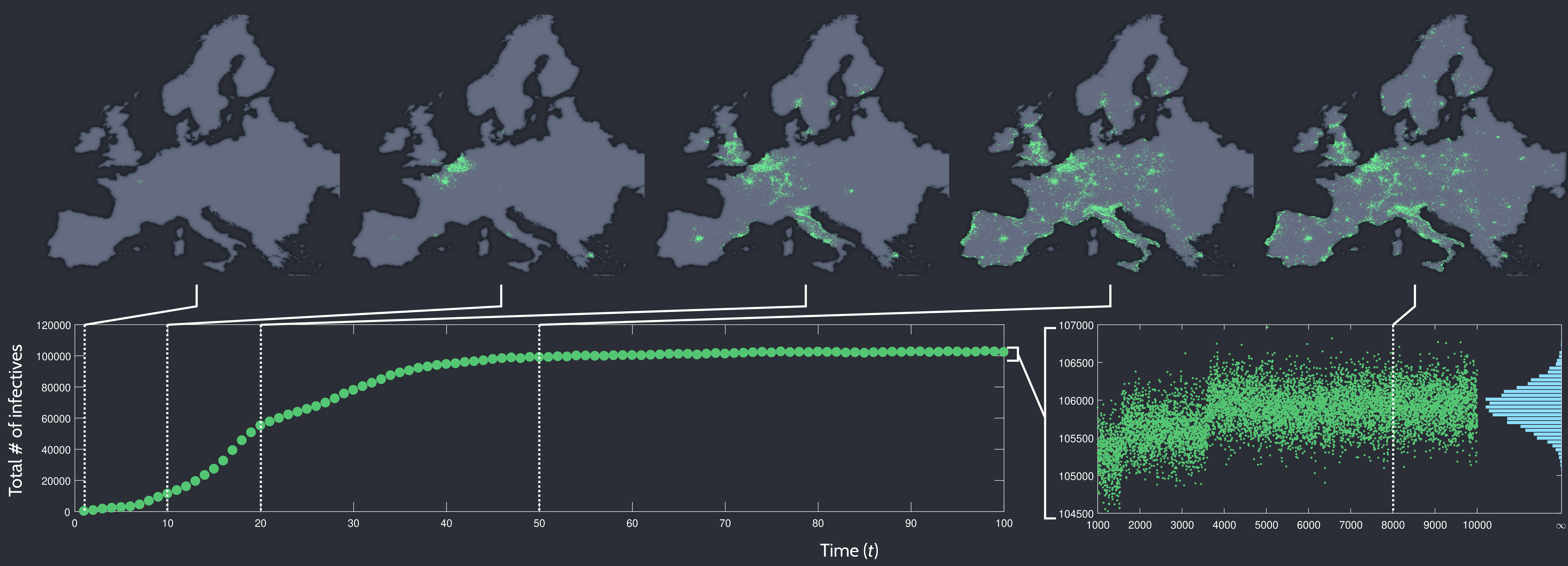

Figure 2: *A simulation of a spatially heterogeneous epidemic model (an example of an occupancy process) using European population density data. The initial outbreak begins in Paris and spreads via airports and local transportation.*

Problem: When there are many nodes, long-term dynamics become complex; classical analysis techniques are ineffective.

Observation: When there are many nodes, trajectories of weighted sums $\sum_{i=1}^n w_i X_{i,t}$ often display Gaussian fluctuations.

Objective: Approximate by Gaussian processes that are simpler to analyse, and use Stein's method to bound the error.

Stein's Method

A general technique to bound the error between $\mathbb{E}g(Y)$ and $\mathbb{E}g(Z)$:

- a random element Z usually approximating Y
- over test functions $g \in \mathcal{G}$.

Idea: Let Y_t be a Markov process with generator L

- starting from the distribution of Y ($Y_0 \sim Y$),
- with stationary distribution $Y_\infty \sim Z$.
- For continuous functions g,

$$\mathbb{E}g(Y) - \mathbb{E}g(Z) = \int_0^\infty \mathbb{E}_Y Lg(Y_t)dt = \mathbb{E}Lf_g(Y),$$

  where $f_g(y) = \int_0^\infty \mathbb{E}_y g(Y_t)dt$.
- Generator L contains information about Z.
- If $Z \sim \mathcal{N}(0,\sigma^2)$ then $Lf(x) = \sigma^2 f(x) - xf(x)$.

The Method:

1. Choose a class of test functions $\mathcal{G}$;
   - define $d_{\mathcal{W}}(X,Z) = \sup_{\|g'\|_\infty \leqslant 1} |\mathbb{E}g(X) - \mathbb{E}g(Z)|$.
2. Bound the functions f_g for each $g \in \mathcal{G}$.
3. Use this with properties of Y to bound $\mathbb{E}Lf_g(Y)$.

Approximations

Estimate evolving probability of occupancy by the dynamical system

$$\boldsymbol{p}_{t+1} = \boldsymbol{P}_t(\boldsymbol{p}_t), \qquad \boldsymbol{p}_0 = \boldsymbol{X}_0$$

about which there is the normal approximation

$$\boldsymbol{Z}_{t+1} = \boldsymbol{p}_{t+1} + D\boldsymbol{P}_t(\boldsymbol{p}_t)(\boldsymbol{Z}_t - \boldsymbol{p}_t) + \mathcal{N}(0, \boldsymbol{p}_t(1-\boldsymbol{p}_t)).$$

Theorem There is a universal constant $c > 0$ such that for any $w \in \mathbb{R}^n$ and $t \geqslant 0$, if $\bar{Z}_t^{(w)} = n^{-1/2}\boldsymbol{w} \cdot \boldsymbol{Z}_t$,

$$d_{\mathcal{W}}\left(\frac{\boldsymbol{w}}{n^{1/2}} \cdot \boldsymbol{X}_t, \ \bar{Z}_t^{(w)}\right) \leqslant \frac{c\kappa_t \|w\|_\infty^3}{\mathrm{Var}(\bar{Z}_t^{(w)})^{3/2}} \sqrt{\frac{1+\log n}{n}}.$$

- κ_t depends on the derivatives of $P_{i,s}$ up to third order, $i = 1,\dots,n$, $s = 0,\dots,t$.
- Error is small if $\partial_j P_{i,t}$ (dependence) is small for each $i,j = 1,\dots,n$, $i \neq j$.

Quasi Stationarity

Ascertaining long-term behaviour is the ultimate application:

- $\boldsymbol{p}_t \to \boldsymbol{p}_\infty$? – *theory of dynamical systems*
- $\boldsymbol{Z}_t \to \boldsymbol{Z}_\infty$? – *standard theory*
- if both, under general assumptions, can show that under some coupling

$$\mathbb{E}\left|\frac{\boldsymbol{w}}{n^{1/2}} \cdot (\boldsymbol{X}_{\tau_n} - \boldsymbol{Z}_\infty)\right| \to 0$$

  where $\tau_n = \mathcal{O}(\log n)$.
- long-term behaviour of $\boldsymbol{X}_t$ is ascertained from that of $\boldsymbol{p}_t$ and $\boldsymbol{Z}_t$.

Acknowledgements

LH is supported by an Australian Postgraduate Award. All authors are supported in part by the Australian Research Council (Discovery Grant DP150101459 and the ARC Centre of Excellence for Mathematical and Statistical Frontiers, CE140100049). Special thanks to Andrew Barbour, Aihua Xia, and Nathan Ross.

ACEMS
AUSTRALIAN RESEARCH COUNCIL CENTRE OF EXCELLENCE FOR MATHEMATICAL AND STATISTICAL FRONTIERS

The University of Queensland
AUSTRALIA