Epidemics with Random Sampling

Amaury Lambert

MCEB 2013, Hameau de l'Étoile, 28 mai 2013

Epidemics with sampling : 2 problems

Assumption :

Each infective is sampled/detected after some random time
(symptoms, medical exam)

Goal :

Infer epidemiological dynamics from...

- **Problem 1.** Pathogen sequence data sampled at all detection times
  = w/ T. Stadler and H. Alexander (ETH Zürich)

- **Problem 2.** Hospital data in antibiotic-resistant epidemics stopped at the first detection time
  = w/ P. Trapman (U. Stockholm)

Epidemics with sampling : 2 problems

Assumption :

Each infective is sampled/detected after some random time (symptoms, medical exam)

Goal :

Infer epidemiological dynamics from...

- **Problem 1.** Pathogen sequence data sampled at **all detection times**
  = w/ T. Stadler and H. Alexander (ETH Zürich)
- **Problem 2.** Hospital data in antibiotic-resistant epidemics stopped **at the first detection time**
  = w/ P. Trapman (U. Stockholm)

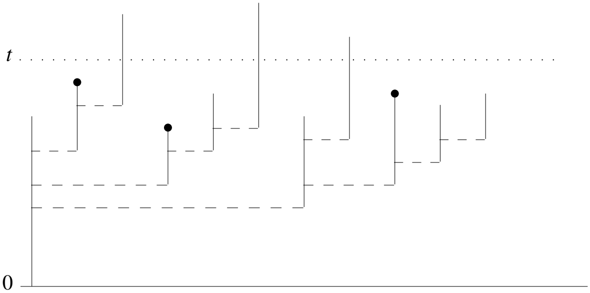
Epidemics with sampling : 2 problems

Assumption :
Each infective is sampled/detected after some random time
(symptoms, medical exam)

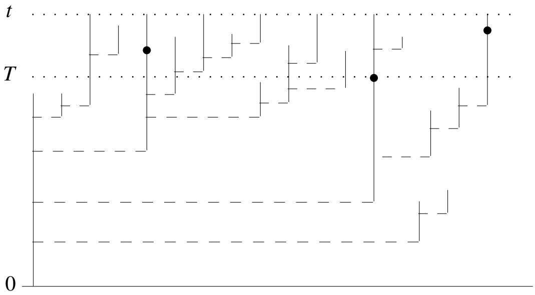
Goal :
Infer epidemiological dynamics from...

- **Problem 1.** Pathogen sequence data sampled at **all detection times**
  = w/ T. Stadler and H. Alexander (ETH Zürich)
- **Problem 2.** Hospital data in antibiotic-resistant epidemics stopped **at the first detection time**
  = w/ P. Trapman (U. Stockholm)

Problem 1 :
Stopping **at all dots before time *t***



Problem 2 :
Stopping **at first dot = time *T***



General framework

- Continuous time
- **Branching process assumption :** Excess of susceptibles
- **Age-dependent** death rate : The duration of infectiousness can have an **arbitrary** distribution
- **Constant** birth rate (transmission)
- **Constant** detection rate

General framework

- Continuous time
- **Branching process assumption :** Excess of susceptibles
- **Age-dependent** death rate : The duration of infectiousness can have an **arbitrary** distribution
- **Constant** birth rate (transmission)
- **Constant** detection rate

General framework

- Continuous time
- **Branching process assumption :** Excess of susceptibles
- **Age-dependent** death rate : The duration of infectiousness can have an **arbitrary** distribution
- **Constant** birth rate (transmission)
- **Constant** detection rate

General framework

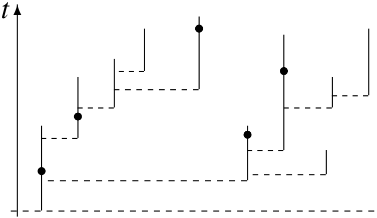
- Continuous time
- **Branching process assumption :** Excess of susceptibles
- **Age-dependent** death rate : The duration of infectiousness can have an **arbitrary** distribution
- **Constant** birth rate (transmission)
- **Constant** detection rate

General framework

- Continuous time
- **Branching process assumption :** Excess of susceptibles
- **Age-dependent** death rate : The duration of infectiousness can have an **arbitrary** distribution
- **Constant** birth rate (transmission)
- **Constant** detection rate

Splitting tree in forward time (Geiger & Kersting 97)

The transmission tree **with sampling** can be described by an asexual population with **marks** where

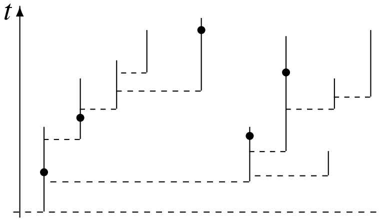


- individuals reproduce independently
- death rate may depend on **age**
- birth rate is constant
- detection rate is constant.

The population size process $(N_t; t \geq 0)$ is a non-Markovian birth–death process.

Splitting tree in forward time (Geiger & Kersting 97)

The transmission tree **with sampling** can be described by an asexual population with **marks** where

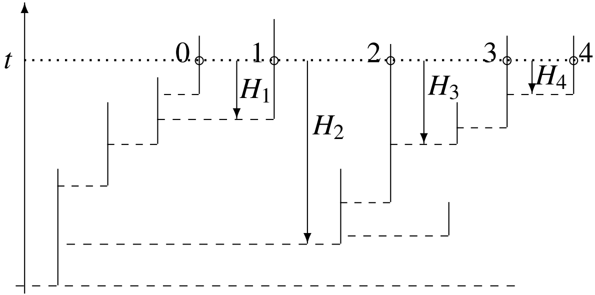


- individuals reproduce independently
- death rate may depend on **age**
- birth rate is constant
- detection rate is constant.

The population size process $(N_t; t \geq 0)$ is a non-Markovian birth–death process.

Without marks (1)

The transmission tree stopped at t can be **oriented** as follows...

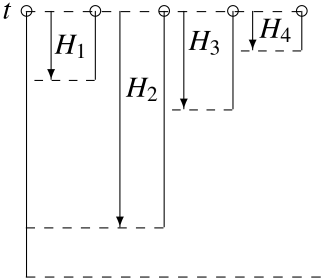


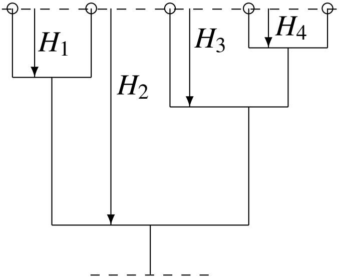
...where the times $H_1, H_2, H_3 \dots$ are the **node depths**.

Without marks (2)

The oriented, **reconstructed tree at *t*** can be represented...
Like this...

...Or like that





Coalescent point process

- **Mathematical result :** The node depths $H_1, H_2, H_3 \dots$ of the tree form a sequence of **independent and identically distributed** positive random variables killed at its first value larger than t
  (Lambert 2010, Lambert & Stadler 2013)

  In particular, conditional on $N_t \neq 0$, the population size N_t follows a geometric distribution.

- Such a tree is called a **coalescent point-process** :
  - fast simulation of reconstructed trees
  - Easy computation of the likelihood of a tree (product form).

Coalescent point process

- **Mathematical result :** The node depths $H_1, H_2, H_3 \dots$ of the tree form a sequence of **independent and identically distributed** positive random variables killed at its first value larger than t

  (Lambert 2010, Lambert & Stadler 2013)

  In particular, conditional on $N_t \neq 0$, the population size N_t follows a geometric distribution.

- Such a tree is called a **coalescent point-process** :
  - fast simulation of reconstructed trees
  - Easy computation of the likelihood of a tree (product form).

Coalescent point process

- **Mathematical result :** The node depths $H_1, H_2, H_3 \ldots$ of the tree form a sequence of **independent and identically distributed** positive random variables killed at its first value larger than t

  (Lambert 2010, Lambert & Stadler 2013)

  In particular, conditional on $N_t \neq 0$, the population size N_t follows a geometric distribution.

- Such a tree is called a **coalescent point-process** :
  - fast simulation of reconstructed trees
  - Easy computation of the likelihood of a tree (product form).

Coalescent point process

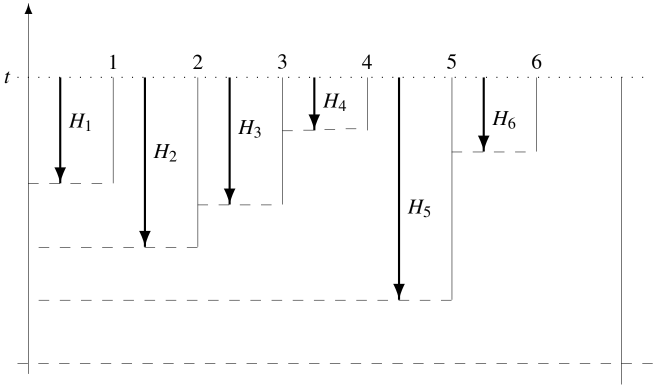
- **Mathematical result :** The node depths $H_1, H_2, H_3 \ldots$ of the tree form a sequence of **independent and identically distributed** positive random variables killed at its first value larger than t

  (Lambert 2010, Lambert & Stadler 2013)

  In particular, conditional on $N_t \neq 0$, the population size N_t follows a geometric distribution.

- Such a tree is called a **coalescent point-process** :
  - fast simulation of reconstructed trees
  - Easy computation of the likelihood of a tree (product form).

Coalescent point process



Likelihood of tree with node depths $(h_i)_{1\le i\le n-1}$

$$\mathscr{L}(\mathscr{T}) = p(t)\prod_{i=1}^{n-1} f(h_i).$$

Adding marks : 2 problems

- **Problem 1.** Likelihood of the tree **spanned by all the distinct detection times ?**
  $\rightsquigarrow$ Previously known : case when death rate is constant (Stadler et al 2012)

- **Problem 2.** Likelihood of the tree **stopped at the first detection time T ?**
  $\rightsquigarrow$ Previously known : N_T is (still) geometric (Trapman & Bootsma 2009)

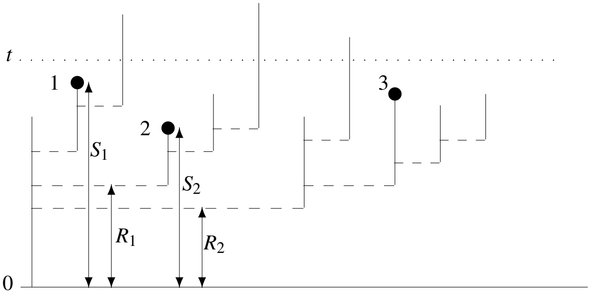
Adding marks : 2 problems

- **Problem 1.** Likelihood of the tree **spanned by all the distinct detection times ?**
  $\rightsquigarrow$ Previously known : case when death rate is constant (Stadler et al 2012)

- **Problem 2.** Likelihood of the tree **stopped at the first detection time T ?**
  $\rightsquigarrow$ Previously known : N_T is (still) geometric (Trapman & Bootsma 2009)

Problem 1 : Assumptions and notation

- A sampled individual immediately leaves the infective population.
- $S_i :=$ sampling time of individual i
- $R_i :=$ coalescence time between individuals $i-1$ and i.



Problem 1 : Result

Theorem (Lambert, Alexander & Stadler 2013)

*The pairs (S_i, R_i) form a **killed Markov chain** with semi-explicit transitions, where the transition probability only depends on the second component (S_i).*

$\leadsto$ Inference of model parameters from viral phylogenies (HIV, flu).
For any given **oriented** tree $\mathscr{T}$ with coalescence times $(x_i)_{2\leq i\leq n}$ and sampling times $(y_i)_{1\leq i\leq n}$,

$$\mathscr{L}(\mathscr{T}) = g(y_1)\,k(y_n)\prod_{i=2}^{n} f(y_{i-1};x_i,y_i).$$

Limitations :

1. Transitions are only semi-explicit ;
2. The Markov chain property depends on the orientation...

Problem 1 : Result

Theorem (Lambert, Alexander & Stadler 2013)

The pairs (S_i, R_i) form a ***killed Markov chain*** *with semi-explicit transitions, where the transition probability only depends on the second component (S_i).*

$\rightsquigarrow$ Inference of model parameters from viral phylogenies (HIV, flu). For any given **oriented** tree $\mathscr{T}$ with coalescence times $(x_i)_{2\leq i\leq n}$ and sampling times $(y_i)_{1\leq i\leq n}$,

$$\mathscr{L}(\mathscr{T}) = g(y_1)\,k(y_n)\prod_{i=2}^{n} f(y_{i-1};x_i,y_i).$$

Limitations :

1. Transitions are only semi-explicit ;
2. The Markov chain property depends on the orientation...

Problem 1 : Result

Theorem (Lambert, Alexander & Stadler 2013)

*The pairs (S_i, R_i) form a **killed Markov chain** with semi-explicit transitions, where the transition probability only depends on the second component (S_i).*

$\rightsquigarrow$ Inference of model parameters from viral phylogenies (HIV, flu). For any given **oriented** tree $\mathscr{T}$ with coalescence times $(x_i)_{2\leq i\leq n}$ and sampling times $(y_i)_{1\leq i\leq n}$,

$$\mathscr{L}(\mathscr{T}) = g(y_1)\,k(y_n)\prod_{i=2}^{n} f(y_{i-1}; x_i, y_i).$$

Limitations :

1. Transitions are only semi-explicit ;
2. The Markov chain property depends on the orientation...

Problem 2 : Assumptions

- Patients have i.i.d lengths of stay in the hospital, all distributed as some r.v. K
- Conditional on infection, the length of stay of a patient is a size-biased version of K
- Detection rate per patient $=: \delta$

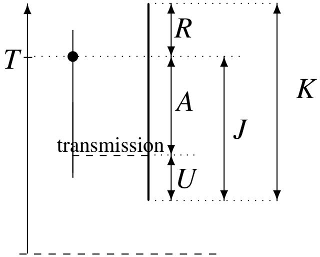
Problem 2 : Notation

For individual i, set

- $U_i :=$ time elapsed from entrance of the hospital up to infection
- $A_i :=$ time elapsed from infection up to T
- $R_i :=$ residual lifetime in the hospital after T.

Set $m := \mathbb{E}(K)$ and let ϕ denote the inverse of the convex function

$$x \mapsto x - \frac{\lambda}{m} \int_{(0,\infty]} (1 - e^{-xy}) \mathbb{P}(K > y)\, dy.$$



Problem 2 : Result

Theorem (Lambert & Trapman 2013)

*Conditional on $N_T = n$, the triples (U_i, A_i, R_i) of the n infectives at time T are **independent and identically distributed**, distributed as*

$$\mathbb{E}(f(U,A,R)) =$$

$$\frac{\lambda}{m}\,\frac{\phi(\delta)}{\phi(\delta)-\delta}\int_{u=0}^{\infty} du \int_{a=0}^{\infty} da \int_{z=u+a}^{\infty} \mathbb{P}(K \in dz)\, e^{-\phi(\delta)a} f(u,a,z-u-a),$$

*In particular, the times $J_i = U_i + A_i$ spent in the hospital up to time T are **independent and identically distributed**, distributed as the r.v. J*

$$\mathbb{P}(J \in dy) = \frac{\lambda/m}{\phi(\delta)-\delta}\,\mathbb{P}(K > y)\left(1 - e^{-\phi(\delta)y}\right) dy.$$

⇝ Inference from hospital data (dates of entrance in the hospital).

Problem 2 : Result

Theorem (Lambert & Trapman 2013)

Conditional on $N_T = n$, the triples (U_i, A_i, R_i) of the n infectives at time T are ***independent and identically distributed****, distributed as*

$$\mathbb{E}(f(U,A,R)) =$$

$$\frac{\lambda}{m}\,\frac{\phi(\delta)}{\phi(\delta)-\delta}\int_{u=0}^{\infty} du \int_{a=0}^{\infty} da \int_{z=u+a}^{\infty} \mathbb{P}(K \in dz)\, e^{-\phi(\delta)a} f(u,a,z-u-a),$$

In particular, the times $J_i = U_i + A_i$ spent in the hospital up to time T are ***independent and identically distributed****, distributed as the r.v. J*

$$\mathbb{P}(J \in dy) = \frac{\lambda/m}{\phi(\delta)-\delta}\,\mathbb{P}(K > y)\left(1 - e^{-\phi(\delta)y}\right) dy.$$

⇝ Inference from hospital data (dates of entrance in the hospital).

Problem 2 : Result

Theorem (Lambert & Trapman 2013)

Conditional on $N_T = n$, the triples (U_i, A_i, R_i) of the n infectives at time T are ***independent and identically distributed****, distributed as*

$$\mathbb{E}(f(U,A,R)) =$$

$$\frac{\lambda}{m}\,\frac{\phi(\delta)}{\phi(\delta)-\delta}\int_{u=0}^{\infty} du \int_{a=0}^{\infty} da \int_{z=u+a}^{\infty} \mathbb{P}(K \in dz)\, e^{-\phi(\delta)a} f(u,a,z-u-a),$$

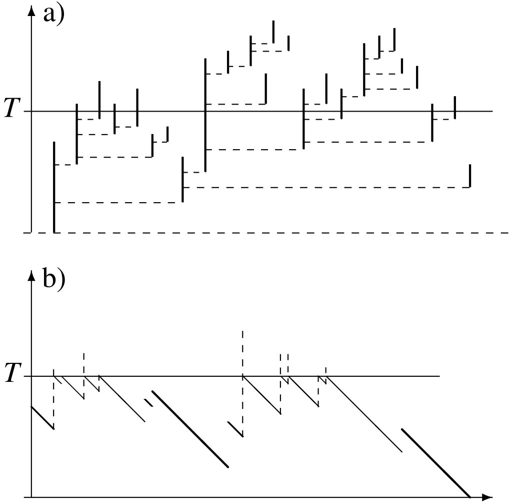
In particular, the times $J_i = U_i + A_i$ spent in the hospital up to time T are ***independent and identically distributed****, distributed as the r.v. J*

$$\mathbb{P}(J \in dy) = \frac{\lambda/m}{\phi(\delta)-\delta}\,\mathbb{P}(K > y)\left(1 - e^{-\phi(\delta)y}\right) dy.$$

$\rightsquigarrow$ Inference from hospital data (dates of entrance in the hospital).

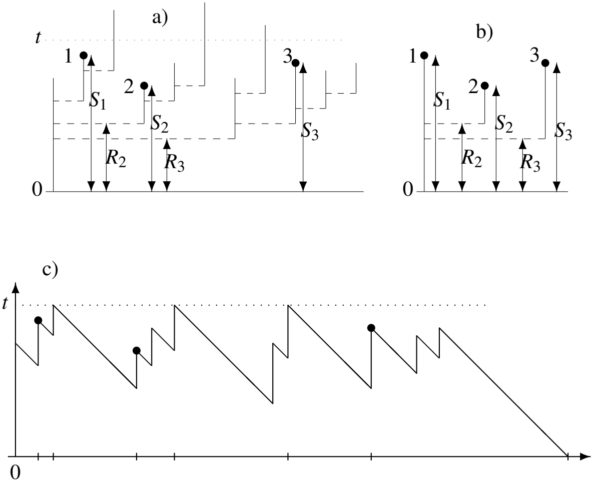
Jumping contour of a tree

a) Binary tree with edge lengths and b) Contour process of its truncation below time t.



Jumping contour of a tree with marks

a) Binary tree with marks, b) its reconstructed tree and c) its contour process.



Co-authors

- ***Helen ALEXANDER*** (ETHZ) . . . .

- ***Tanja STADLER*** (ETHZ) . . . .

- ***Pieter TRAPMAN*** (U. Stockholm) . . . .

SMILE : A cross-disciplinary group in CIRB

- CIRB = Center for Interdisciplinary Research in Biology (Collège de France)
- SMILE = Stochastic Models for the Inference of Life Evolution

SMILE group in June 2012

Institutions

- ***Stochastic Models for the Inference of Life Evolution* (SMILE)**

  ⊂ Center for Interdisciplinary Research in Biology

  ⊂ Collège de France

  CIRB COLLÈGE DE FRANCE 1530

- ***Stochastics & Biology group***

  ⊂ Laboratoire de Probabilités et Modèles Aléatoires

  ⊂ UPMC University Paris 06

  LPMA UPMC SORBONNE UNIVERSITÉS

- **ANR *Modèles Aléatoires eN Écologie, Génétique, Évolution* (MANEGE)**

  AGENCE NATIONALE DE LA RECHERCHE ANR

Conference announcement

Mathematics for an Evolving Biodiversity

Montréal, Canada

September 16–20, 2013

Organizers : Jonathan Davies (McGill), Nicolas Lartillot (CNRS & U. Montréal) and myself

http://www.crm.umontreal.ca/2013/Biodiversity13/