Markov models for accumulating mutations

BY N. BEERENWINKEL

Department of Biosystems Science and Engineering, ETH Zurich,
Mattenstrasse 26, 4058 Basel, Switzerland
niko.beerenwinkel@bsse.ethz.ch

AND S. SULLIVANT

Department of Mathematics, North Carolina State University,
Raleigh, North Carolina 27607, U.S.A.
smsulli2@ncsu.edu

SUMMARY

We introduce and analyze a waiting time model for the accumulation of genetic changes. The continuous-time conjunctive Bayesian network is defined by a partially ordered set of mutations and by the rate of fixation of each mutation. The partial order encodes constraints on the order in which mutations can fixate in the population, shedding light on the mutational pathways underlying the evolutionary process. We study a censored version of the model and derive equations for an EM algorithm to perform maximum likelihood estimation of the model parameters. We also show how to select the maximum likelihood partially ordered set. The model is applied to genetic data from cancer cells and from drug resistant human immunodeficiency viruses, indicating implications for diagnosis and treatment.

Some key words: Bayesian network; Cancer; Genetic progression; HIV; Partially ordered set; Poset.

1. INTRODUCTION

Pathogen evolution is characterized by recurrent phases of evolutionary escape from opposing selective pressure. During these stages several mutations are available to the population that increase fitness. Adaptive evolution proceeds by accumulating these beneficial genetic changes. For example, in HIV infection, the virus acquires mutations in cytotoxic T lymphocyte epitopes that interfere with the human immune response, while drug resistant mutations develop under antiviral treatment. Similarly, the genetic progression of cancer is characterized by the accumulation of mutations in oncogenes and tumor suppressor genes, which confer a selective advantage.

Evolutionary escape dynamics have been analyzed from a population genetics perspective. This approach clarified the influence of basic parameters such as population size and mutation rate on the risk of escape. It also highlighted the importance of the topology of escape, i.e. the structure of the possible mutational pathways (Iwasa et al., 2003; Beerenwinkel et al., 2006).

Here we introduce a model for the accumulation of genetic changes that allows for inference about the topology and the speed of evolutionary escape. The continuous-time conjunctive Bayesian network is a continuous-time Markov chain model, defined by a partially ordered set of advantageous mutations and their rates of fixation. The partial order encodes constraints on the succession in which mutations can occur and fixate in the population. We assume that the fixation

times follow independent exponential distributions. The waiting process for a mutation starts only when all of its predecessor mutations have occurred. The order constraints and waiting times reveal important information on the underlying biological process with implications for diagnosis and treatment. We exemplify the use of these Bayesian network models with the development of drug resistance in HIV and the genetic progression of prostate cancer.

The continuous-time conjunctive Bayesian network is a continuous-time analogue of the discrete conjunctive Bayesian network introduced by Beerenwinkel et al. (2006), which was shown to have very desirable statistical and algebraic properties (Beerenwinkel et al., 2007). Both networks are natural exponential families with closed-form expressions for maximum likelihood estimation and model selection. We argue that the continuous-time network is the more natural model for accumulating mutations, and we explore the connection to the discrete network.

One difference between the discrete and the continuous-time model is the state space. A state for the continuous-time model is a vector of time-points of the mutational events, while a state for the discrete model is a vector of binary indicators of the mutations. Despite the biological realism of the continuous-time model, the observed data rarely consist of the time-points of the genetic changes. Instead, we typically observe discrete genotypes at randomly sampled time-points. To address this discrepancy, we study a censored version of the continuous-time conjunctive Bayesian network and we derive an iterative algorithm for maximum likelihood estimation.

A special case of the discrete model, where the partially ordered set is a tree, is known as the oncogenetic or mutagenetic tree model (Desper et al., 1999; Beerenwinkel et al. 2005b, 2005c; Beerenwinkel & Drton, 2007). It has been applied to the somatic evolution of cancer (Radmacher et al., 2001; Rahnenführer et al., 2005) and to the evolution of drug resistance in HIV (Beerenwinkel et al., 2005a). A related tree model by von Heydebreck et al. (2004) represents the genetic changes at the leaves of the tree and regards the interior vertices as hidden events. Several authors have considered larger model classes, including general Bayesian networks (Simon et al., 2000; Deforche et al., 2006) and more general Markov chain models (Foulkes & DeGruttola, 2003; Hjelm et al., 2006). As compared to trees and partially ordered sets, these models are more flexible in describing mutational pathways, but parameter estimation and model selection is considerably more difficult. In fact, the number of free parameters of these models is typically exponential in the number of mutations, whereas in the continuous-time conjunctive Bayesian network, it is equal to the number of mutations. We demonstrate that parameter estimation and selection of an optimal partially ordered set can be performed efficiently for continuous-time conjunctive Bayesian networks which thus provide an attractive framework for modelling the accumulation of mutations, even if the number of mutations is moderate or large.

2. CONTINUOUS-TIME CONJUNCTIVE BAYESIAN NETWORKS

In this section, we introduce and describe some of the basic properties of continuous-time conjunctive Bayesian networks. These models are continuous-time Markov chain models on the distributive lattice of a partially ordered set. We first review some background material from combinatorics; see also the introductory sections of Beerenwinkel et al. (2006) or Stanley (1999).

A partially ordered set P has a binary relation $\prec$, which is reflexive, antisymmetric and transitive. In our models, the set P will be a set of genetic events, and the relation specifies constraints on the order in which they occur: $p \prec q$ means that event p happens before or at the same time as event q. We assume throughout that events that occur simultaneously have been collapsed in a pre-processing step. We say that p is a parent of q if $p \prec q$ and there is no element $r \in P$ such that $r \neq p$, q and $p \prec r \prec q$. We use the notation $p \to q$ to denote that p is a parent of q and call $p \to q$ a cover relation. The set of all parents of an element q is denoted

Partially ordered set, P Waiting times, T Lattice of order ideals, $J(P)$

$$Z_i \sim \mathrm{Exp}(\lambda_i) \ \ (i = 1, 2, 3, 4)$$

$$T_1 = Z_1$$
$$T_2 = Z_2$$
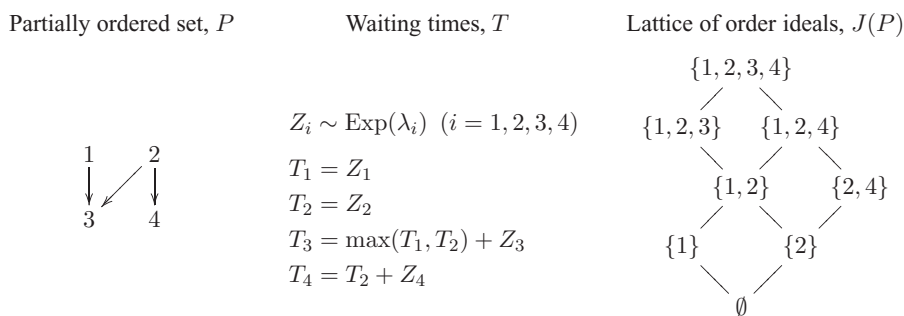$$T_3 = \max(T_1, T_2) + Z_3$$
$$T_4 = T_2 + Z_4$$

Fig. 1. Running example of a continuous-time conjunctive Bayesian network model.

by $\mathrm{pa}(q)$. The set of cover relations is used to visualize a partially ordered set by means of its Hasse diagram.

The distributive lattice of order ideals of P, denoted by $J(P)$, consists of all subsets $S \subseteq P$, that are closed downward, i.e. $S \in J(P)$ if and only if, for all $q \in S$ and $p \prec q$, we have that $p \in S$. The order ideals of P correspond to the genotypes, or mutational patterns, that are compatible with the order constraints. We refer to $\emptyset \in J(P)$ as the wild type.

Example 1. As a running example, consider the partially ordered set P on the four element set $[4] = \{1, 2, 3, 4\}$ subject to the order relations $1 \prec 3$, $2 \prec 3$, $2 \prec 4$. The distributive lattice $J(P)$ of order ideals of P consists of eight elements. Both sets are displayed in Fig. 1.

Let P be a partially ordered set that equals $[n]$ as a set. For each event $i \in P$, we define a random variable $Z_i \sim \mathrm{Exp}(\lambda_i)$. Then we define the random variables $T_i = \max_{j \in \mathrm{pa}(i)} T_j + Z_i$ $(i = 1, \ldots, n)$. The random variable T_i describes how long we have to wait until event i occurs, assuming that we start at time zero with no events. Mutation i cannot occur until all the mutations preceding it in the partial order P have occurred. The family of joint distributions of T defined in this manner is the continuous-time conjunctive Bayesian network. It has state space $\mathbb{R}^n_{>0}$ consisting of vectors of waiting times and parameters $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n_{>0}$.

The family of conjunctive Bayesian networks is not an a priori Bayesian family of models. The terminology Bayesian network comes from the graphical models community. For brevity, we will often refer to the continuous-time conjunctive Bayesian network as the continuous-time network, and to the discrete conjunctive Bayesian network as the discrete network.

The recursive definition of the discrete model implies that its probability density function is

$$f_{P,\lambda}(t) = \begin{cases} \prod_{i=1}^n \lambda_i \exp\{-\lambda_i(t_i - \max_{j \in \mathrm{pa}(i)} t_j)\}, & t_i > \max_{j \in \mathrm{pa}(i)} t_j, \ i \in [n], \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

The continuous-time network is a regular exponential family, with minimal sufficient statistic consisting of the vector of time differences $(t_i - \max_{j \in \mathrm{pa}(i)} t_j)_{i \in [n]}$.

One instance of the random variable T is a vector of time-points $t = (t_1, \ldots, t_n)$ that satisfy the inequality relations implied by P, i.e. $t_i > \max_{j \in \mathrm{pa}(i)} t_j$ for all $i \in [n]$. A set of time-points satisfying these constraints is compatible with the partially ordered set. The data for the model is a matrix of time-points $(t_{ki})_{k \in [N], \, i \in [n]}$, where N is the number of observations.

PROPOSITION 1. *Let P be a partially ordered set and (t_{ki}) be a collection of data. If any of the observations t_k. are incompatible with P, then the likelihood function is identical to zero.*

Otherwise, the maximum likelihood estimate of λ is

$$\hat{\lambda}_i = \frac{N}{\sum_{k=1}^{N}(t_{ki} - \max_{j \in \text{pa}(i)} t_{kj})}, \quad i \in [n].$$

Proof. If $t_{1\cdot}, \ldots, t_{N\cdot}$ are compatible with P, then the loglikelihood function is

$$\ell(\lambda_1, \ldots, \lambda_n) = \sum_{k=1}^{N} \sum_{i=1}^{n} \left\{ \log \lambda_i - \lambda_i \left(t_{ki} - \max_{j \in \text{pa}(i)} t_{kj} \right) \right\}.$$

Differentiating with respect to λ_i yields the equations

$$\sum_{k=1}^{N} \left\{ \frac{1}{\lambda_i} - \left(t_{ki} - \max_{j \in \text{pa}(i)} t_{kj} \right) \right\} = 0,$$

and the claimed formula follows by solving for λ_i. □

THEOREM 1. *Given data (t_{ki}), the maximum likelihood partially ordered set is the largest partially ordered set that is compatible with the data.*

Proof. Suppose that the data is compatible with two partially ordered sets $(P^1, \prec_1)$ and $(P^2, \prec_2)$, and that P^1 is a refinement of P^2 (i.e. every relation that holds in P^2 also holds in P^1). We show that the likelihood function at the maximum likelihood estimate is larger for P^1 than for P^2. This implies that adding relations compatible with the data increases the likelihood.

According to Proposition 1, the maximum likelihood estimates $\hat{\lambda}^1$ for P^1 and $\hat{\lambda}^2$ for P^2 are

$$\hat{\lambda}_i^l = \frac{N}{\sum_{k=1}^{N}(t_{ki} - \max_{j \in \text{pa}_l(i)} t_{kj})}, \quad i \in [n], \quad l \in [2].$$

We can replace $\text{pa}_l(i)$ with the set $\{j \in P^l \mid j \prec_l i\}$. Since P^1 has more relations than P^2, this implies that the maximum is taken over a strictly larger set, and thus $\hat{\lambda}_i^1 \geqslant \hat{\lambda}_i^2$ for all i.

However, the loglikelihood function evaluated at $\hat{\lambda}^l$ is

$$\ell_l(\hat{\lambda}^l) = \sum_{k=1}^{N} \sum_{i=1}^{n} \left\{ \log \hat{\lambda}_i^l - \hat{\lambda}_i^l \left(t_{ki} - \max_{j \in \text{pa}_l(i)} t_{kj} \right) \right\}$$

$$= \sum_{i=1}^{n} \left\{ N \log \hat{\lambda}_i^l - \hat{\lambda}_i^l \sum_{k=1}^{N} \left(t_{ki} - \max_{j \in \text{pa}_l(i)} t_{kj} \right) \right\}$$

$$= \sum_{i=1}^{n} \left(N \log \hat{\lambda}_i^l - N \right).$$

Since the logarithm is a monotone function, we deduce that $\ell_1(\hat{\lambda}^1) \geqslant \ell_2(\hat{\lambda}^2)$. □

One of the most interesting quantities we can compute using the continuous-time network is the expected waiting time until a particular pattern $S \in J(P)$ is reached. Assuming that the parameters λ are known, we are asking how long it takes until a set of genetic events have occurred. The expected waiting time is an important measure of genetic progression. Rahnenführer et al. (2005) have shown that it is a prognostic factor of survival and time to relapse in glioblastoma and prostate cancer patients, respectively, even after adjustment for traditional clinical markers.

Since the exponential distribution is memoryless, calculating the waiting time from the wild type will also determine the waiting time between any two patterns. Furthermore, the nature of

the conditional factorization for the joint density of T implies that we can restrict attention to the case where $S = P$, i.e. to determining the waiting time until all events have occurred.

Let $S \in J(P)$ be an observable genotype. We define $\mathrm{Exit}(S) = \{j \in P \mid j \notin S, S \cup \{j\} \in J(P)\}$ to be the set of events that have not occurred in S, but could occur next. For any subset $T \subseteq P$, we set $\lambda_T = \sum_{j \in T} \lambda_j$. A chain in the distributive lattice $J(P)$ is a collection of subsets $C_0, C_1, \ldots, C_k \in J(P)$ that satisfy $C_i \subset C_{i+1}$ and $C_i \neq C_{i+1}$, for all i. A chain is maximal, if it is as long as possible. All maximal chains in the distributive lattice $J(P)$ have length $n + 1$ with $n = |P|$ and start with $C_0 = \emptyset$ and end with $C_n = P$. Let $\mathcal{C}\{J(P)\}$ denote the collection of maximal chains in $J(P)$; a typical element is denoted by $C = (C_0, \ldots, C_n)$.

THEOREM 2. *The expected waiting time until all events have occurred is*

$$E\left(\max_{i \in P} T_i\right) = \lambda_1 \cdots \lambda_n \sum_{C \in \mathcal{C}\{J(P)\}} \left(\prod_{i=0}^{n-1} \frac{1}{\lambda_{\mathrm{Exit}(C_i)}}\right) \left(\sum_{i=0}^{n-1} \frac{1}{\lambda_{\mathrm{Exit}(C_i)}}\right).$$

Before proving Theorem 2, we sketch the idea of the proof, which is a common technique for proofs throughout the paper. The indicated expectation involves the integral of a function that depends on maxima, which are not simple to integrate directly. We first decompose the integral into a sum of integrals over many different regions, one for each maximal chain in $J(P)$. Over these simpler regions, the maximum function disappears. Furthermore, these regions are each simplicial cones and the integral can then be computed by a simple change of coordinates.

Proof. Let $f(t)$ be the density function from equation (1). We must compute

$$\int_{\mathbb{R}^n_{\geq 0}} \max_{i \in P} t_i f(t) \, dt. \tag{2}$$

Let S_n denote the symmetric group on n letters with $\sigma = (\sigma_1, \ldots, \sigma_n)$ a typical element. The integral (2) over the positive orthant breaks up as the sum

$$\sum_{\sigma \in S_n} \int_{t_{\sigma_1}=0}^{\infty} \int_{t_{\sigma_2}=t_{\sigma_1}}^{\infty} \cdots \int_{t_{\sigma_n}=t_{\sigma_{n-1}}}^{\infty} t_{\sigma_n} f(t) dt.$$

That is, the sum breaks up the integral into smaller integrals over regions $0 < t_{\sigma_1} < \cdots < t_{\sigma_n}$. The integrand is zero unless $\sigma_1, \ldots, \sigma_n$ is a linear extension of P. In other words, the integrand is zero unless the sets $C_i = \cup_{j=1}^{i} \{\sigma_j\}$ for $i = 0, \ldots, n$ form a maximal chain in the distributive lattice $J(P)$. So suppose that σ is a linear extension of P. Without loss of generality, we may suppose that this linear extension is $1 \prec \cdots \prec n$. We must compute the integral

$$\int_{t_1=0}^{\infty} \int_{t_2=t_1}^{\infty} \cdots \int_{t_n=t_{n-1}}^{\infty} t_n f(t) \, dt,$$

where, over this restricted region, $f(t)$ now has the form

$$f(t) = \prod_{i=1}^{n} \lambda_i \exp\{-\lambda_i(t_i - t_{j(i)})\},$$

where $j(i)$ is the largest number with $j(i) \prec i$ in P.

Now introduce the change of coordinates $u_0 = t_1$, $u_i = t_{i+1} - t_i$, $i \in [n-1]$. The determinant of this linear transformation is one, so the integral becomes

$$\int_{u_0=0}^{\infty} \cdots \int_{u_{n-1}=0}^{\infty} (u_0 + \cdots + u_{n-1}) \prod_{i=1}^{n} \lambda_i \exp\{-\lambda_i(u_{i-1} + u_{i-2} + \cdots + u_{j(i)})\} du. \tag{3}$$

The multiple integral is now over a product domain, and involves a function in product form, so we want to break (3) into the product of integrals. To do this, we must collect the λ_i terms that go with the various u_k terms. In the exponent, we have that λ_i appears as a coefficient of u_k if and only if $i > k \geqslant j(i)$. This, in turn, implies that when all the events $1, \ldots, k$ have occurred, all the predecessor events of i have occurred. This means that $i \in \text{Exit}(C_k)$, where $C_k = \{1, \ldots, k\}$. Thus, the transformed integral breaks up as a sum of n integrals that have the form:

$$\lambda_1 \cdots \lambda_n \int_{u_0=0}^{\infty} \cdots \int_{u_{n-1}=0}^{\infty} u_j \prod_{i=0}^{n-1} \exp\left\{ -\lambda_{\text{Exit}(C_i)} u_i \right\} du.$$

The integral is over a product domain of a product function, so equals

$$\lambda_1 \cdots \lambda_n \frac{1}{\lambda_{\text{Exit}(C_j)}} \prod_{i=0}^{n-1} \frac{1}{\lambda_{\text{Exit}(C_i)}},$$

which completes the proof.   □

Theorem 2 yields a formula for the expected waiting time until all mutations have occurred, which, at first glance seems to require the enumeration of all maximal chains in the distributive lattice $J(P)$. However, this sum can be computed by a dynamic programming algorithm.

PROPOSITION 2. *For each $S \in J(P)$ define P_S and E_S by*

$$P_S = \sum_{j \in S: S \setminus \{j\} \in J(P)} \frac{\lambda_j}{\lambda_{\text{Exit}(S \setminus \{j\})}} P_{S \setminus \{j\}},$$

$$E_S = \sum_{j \in S: S \setminus \{j\} \in J(P)} \left(\frac{\lambda_j}{\lambda_{\text{Exit}(S \setminus \{j\})}} E_{S \setminus \{j\}} + \frac{\lambda_j}{\lambda_{\text{Exit}(S \setminus \{j\})}^2} P_{S \setminus \{j\}} \right),$$

subject to the initial conditions $P_\emptyset = 1$ and $E_\emptyset^i = 0$. Then $E(\max_{i \in P} T_i) = E_{[n]}$.

We omit the proof of Proposition 2, which closely follows the proof of Proposition 4 below.

3. Relation to the discrete conjunctive Bayesian network

In this section, we explore the connection between the continuous-time and the discrete network (Beerenwinkel et al., 2007). We want to understand how the models relate to each other and how structural information from one model yields information about the other. We are naturally led to study discrete models because we rarely have access to the times at which the individual events occurred, but can only observe which events have occurred after a certain sampling time.

We will show that the discrete network gives a first-order approximation to the transition probabilities in the continuous-time network. This suggests that the discrete network is not optimal from a modelling standpoint as the nature of our applications is to wait until mutations occur. However, the discrete model is much simpler to work with, and its maximum likelihood estimates can be used as an initial step in iterative algorithms for maximum likelihood estimation in the censored versions of the continuous-time network described in § 4.

To derive this result, we consider the continuous-time network as a continuous-time Markov chain on the distributive lattice $J(P)$. See Norris (1997) for background on Markov chains. The

rate matrix for the Markov chain is the $m \times m$ matrix Q, where $m = |J(P)|$, defined by

$$Q_{S,T} = \begin{cases} \lambda_j, & S \subset T, T \backslash S = \{j\}, \\ -\lambda_{\text{Exit}(S)}, & S = T, \\ 0, & \text{otherwise}, \end{cases}$$

for all $S, T \in J(P)$. If we fix a linear extension of $J(P)$ and order the rows and columns of Q according to this linear extension, Q will be an upper triangular matrix.

Let $p(t)$ be the $m \times m$ matrix where, for $S, T \in J(P)$, the entry $p_{S,T}(t)$ denotes the probability that the continuous-time Markov chain with state space $J(P)$, is in state T at time t starting from state S at time 0. This quantity can be calculated by integrating the density function from the continuous-time model. However, it is simpler to use standard theory of Markov chains. The matrix $p(t)$ is the solution of the system of differential equations $dp(t)/dt = Qp(t)$ subject to the initial conditions $p(0) = I$, the identity matrix. The solution to this first-order differential equation is the matrix exponential $p(t) = \exp(Qt)$. It satisfies $d^k p(t)/dt^k \mid_{t=0} = Q^k$. So a first-order approximation to $p(t)$ is a function $\tilde{p}(t)$ that satisfies $\tilde{p}(0) = I$ and $d\tilde{p}(t)/dt \mid_{t=0} = Q$.

The first-order approximation can be derived as follows. Associated to the discrete network are n parameters $\theta_1, \ldots, \theta_n$, where θ_i is the conditional probability that event i has occurred, given that all its predecessor events have occurred. By setting $\theta_i = 1 - \exp(-\lambda_i t)$ and defining

$$\tilde{p}_{S,T}(t) = \begin{cases} \prod_{i \in T \backslash S} \theta_i \prod_{i \in \text{Exit}(T)} (1 - \theta_i), & S \subseteq T, \\ 0, & \text{otherwise}, \end{cases}$$

the discrete model is naturally interpreted as a continuous-time model, where $\tilde{p}_{S,T}(t)$ is the probability that the discrete model is in state T at time t given an initial state S at time 0.

PROPOSITION 3. *The model $\tilde{p}(t)$ derived from the discrete conjunctive Bayesian network is a first-order approximation to the continuous-time conjunctive Bayesian network $p(t)$.*

Proof. Diagonal entries of $\tilde{p}(t)$ have the form $\prod \exp(-\lambda_i t)$, and off-diagonal entries are either identical to zero or are a product of terms, at least one of which is of the form $1 - \exp(-\lambda_i t)$. This implies that $\tilde{p}(0) = I$.

If $T = S$, then $\tilde{p}_{S,S}(t) = \prod_{i \in \text{Exit}(S)} \exp(-\lambda_i t)$. So, by the product rule

$$\frac{d}{dt} \tilde{p}_{S,S}(t) = \sum_{i \in \text{Exit}(S)} -\lambda_i \tilde{p}_{S,S}(t)$$

and thus $d\tilde{p}_{S,S}(t)/dt \mid_{t=0} = -\lambda_{\text{Exit}(S)} = Q_{S,S}$. If $S \subset T$ then

$$\tilde{p}_{S,T}(t) = \prod_{i \in T \backslash S} \{1 - \exp(-\lambda_i t)\} \prod_{i \in \text{Exit}(T)} \exp(-\lambda_i t)$$

and thus

$$\frac{d}{dt} \tilde{p}_{S,T}(t) = \sum_{i \in T \backslash S} \lambda_i \frac{\exp(-\lambda_i t)}{1 - \exp(-\lambda_i t)} \tilde{p}_{S,T}(t) - \sum_{i \in \text{Exit}(T)} \lambda_i \tilde{p}_{S,T}(t).$$

If $T \backslash S$ has cardinality greater than one, then $d\tilde{p}_{S,T}(t)/dt \mid_{t=0} = 0 = Q_{S,T}$, because every term in the sum involves at least one expression of the form $1 - \exp(-\lambda_i t)$. On the other hand, if $T = S \cup \{j\}$, then $d\tilde{p}_{S,T}(t)/dt \mid_{t=0} = \lambda_j = Q_{S,T}$, because only the first term in the sum does not contain an expression of the form $1 - \exp(-\lambda_i t)$. As all other entries in $\tilde{p}(t)$ and Q are zero, this proves that $\tilde{p}(t)$ is a first-order approximation to $p(t)$. $\qquad\square$

Given that the discrete model is a first-order approximation to the continuous-time model, it seems natural to conjecture that these two models are equal. If P is the partially ordered set with no relations, then the two models coincide. However, if P contains at least one relation, the models are no longer the same, and the discrete network is not even a second-order approximation to the continuous-time network. This is illustrated in the following example.

Example 2. Let P be the partially ordered set on two elements with one relation $1 \prec 2$ and fix the natural order $\emptyset, \{1\}, \{1, 2\}$ in $J(P)$. If $\lambda_1 \neq \lambda_2$, then

$$\tilde{p}(t) = \left\{ \begin{array}{ccc} e^{-\lambda_1 t} & (1 - e^{-\lambda_1 t})e^{-\lambda_2 t} & (1 - e^{-\lambda_1 t})(1 - e^{-\lambda_2 t}) \\ 0 & e^{-\lambda_2 t} & 1 - e^{-\lambda_2 t} \\ 0 & 0 & 1 \end{array} \right\},$$

$$p(t) = \left\{ \begin{array}{ccc} e^{-\lambda_1 t} & \{\lambda_1/(\lambda_1 - \lambda_2)\}(e^{-\lambda_2 t} - e^{-\lambda_1 t}) & 1 - (\lambda_1 e^{-\lambda_2 t} - \lambda_2 e^{-\lambda_1 t})/(\lambda_1 - \lambda_2) \\ 0 & e^{-\lambda_2 t} & 1 - e^{-\lambda_2 t} \\ 0 & 0 & 1 \end{array} \right\}.$$

In particular, $d^2 \tilde{p}_{\emptyset,\{1\}}(t)/dt^2 \mid_{t=0} = -\lambda_1^2 - 2\lambda_1\lambda_2$, whereas $(Q^2)_{\emptyset,\{1\}} = -\lambda_1^2 - \lambda_1\lambda_2$.

The discrepancies exhibited in this example become more dramatic as P develops longer chains. While the discrete network is not identical to the continuous-time network, its nice properties can be exploited at various points during optimization.

4. CENSORING

In this section, we define and analyze the censored continuous-time conjunctive Bayesian network model that we will apply to genetic data in §5. The reason for introducing censoring is that we typically do not know the time-points $t_1, \ldots, t_n$ at which the events have occurred. Often we can only measure, at a particular time, which of the events have occurred so far. It is natural to assume that the observation times are also random. For example, HIV drug resistance mutations can only be detected after therapy failure.

We introduce a new event s for the sampling process, such that the random variable T_s is an independent, exponentially distributed waiting time (the sampling time, or observation time), $T_s \sim \text{Exp}(\lambda_s)$. We define a new partially ordered set $P_s = P \cup \{s\}$ by adding the element s with no relations to the other elements in P. An observed mutational pattern $S \subseteq [n]$ results from the event $T_i < T_s$ for all $i \in S$ and $T_s < T_i$ for all $i \in [n] \setminus S$.

An observed set of events, S, imposes extra relations $i \prec s$ for $i \in S$ and $s \prec i$ for $i \in [n] \setminus S$ on the partially ordered set P_s, and we are led to study this refined poset Q_S.

A realization of the random vector T is said to be compatible with Q, denoted by $T \vdash Q$, if $T_i < T_j$ whenever $i \prec j$ in Q. We can directly compute the probability of the event $T \vdash Q$ in terms of the distributive lattices $J(Q)$ and $J(P_s)$. Throughout this section, we abuse notation and say that $T \sim P_s$ if $T = (T_s, T_1, \ldots, T_n)$ is distributed according to the continuous-time conjunctive Bayesian network associated to P_s with parameter vector $\lambda = (\lambda_s, \lambda_1, \ldots, \lambda_n)$.

THEOREM 3. *The probability that $T \sim P_s$ is compatible with the partially ordered set Q is*

$$\text{pr}(T \vdash Q) = \lambda_s \lambda_1 \cdots \lambda_n \sum_{C \in \mathcal{C}(J(Q))} \prod_{i=0}^{n} \frac{1}{\lambda_{\text{Exit}(C_i)}}, \tag{4}$$

where the sum runs over all maximal chains in the distributive lattice $J(Q)$.

Remark 1. In this formula, and in all formulae throughout this section, the expression $\text{Exit}(S)$ always refers to the underlying partially ordered set P_s and not to the refinement Q.

Proof of Theorem 3. We must compute the integral

$$\int_{\mathbb{R}^{n+1}_{\geqslant 0}} \mathbb{I}_Q(t)f(t)dt,$$

where $\mathbb{I}_Q(t)$ is the indicator function of compatibility with Q. The integral breaks up into a sum over the linear extensions of Q over regions over the form $t_{\sigma_0} < \cdots < t_{\sigma_n}$. Without loss of generality and after renaming the elements of P_s, we can assume that the linear extension of interest is $0 \prec \cdots \prec n$. We must calculate the integral

$$\int_{t_0=0}^{\infty} \int_{t_1=t_0}^{\infty} \cdots \int_{t_n=t_{n-1}}^{\infty} f(t)dt,$$

where over the restricted region, the integrand has the form

$$f(t) = \prod_{i=0}^{n} \lambda_i \exp\{-\lambda_i(t_i - t_{j(i)})\}.$$

With the change of variables $u_0 = t_0$ and $u_{i+1} = t_{i+1} - t_i$, $i \in [n]$, the integral becomes

$$\prod_{i=0}^{n} \int_{u_i=0}^{\infty} \exp(-\lambda_{\text{Exit}(C_i)}u_i)du_i,$$

where $C_i = \{0, 1, \ldots, i-1\}$. This yields the desired contribution to the integral. □

Example 3. Let P be the partially ordered set from Example 1 with relations $1 \prec 3, 2 \prec 3$ and $2 \prec 4$, and consider the partially ordered set $P_s = P \cup \{s\}$ with no additional relations. Suppose we want to calculate the probability of mutations 2 and 4 occurring before measurement. The refinement $Q_{2,4}$ corresponding to the genotype $\{2, 4\}$ is a chain $2 \prec 4 \prec s \prec 1 \prec 3$, and so the distributive lattice $J(Q_{2,4})$ is also a chain. From equation (4), the probability of $T \vdash Q_{2,4}$ is

$$\lambda_1\lambda_2\lambda_3\lambda_4\lambda_s \frac{1}{\lambda_1 + \lambda_2 + \lambda_s} \frac{1}{\lambda_1 + \lambda_4 + \lambda_s} \frac{1}{\lambda_1 + \lambda_s} \frac{1}{\lambda_1} \frac{1}{\lambda_3}.$$

On the other hand, the distributive lattice $J(Q_{1,2})$ has four chains and $\text{pr}(T \vdash Q_{1,2})$ is the sum of four terms of product form. The expression can be written as

$$\frac{\lambda_1\lambda_2\lambda_3\lambda_4\lambda_s}{(\lambda_1 + \lambda_2 + \lambda_s)(\lambda_3 + \lambda_4 + \lambda_s)(\lambda_3 + \lambda_4)} \left(\frac{1}{\lambda_2 + \lambda_s} + \frac{1}{\lambda_1 + \lambda_4 + \lambda_s}\right) \left(\frac{1}{\lambda_3} + \frac{1}{\lambda_4}\right).$$

A general recursive formula for this probability is given in Proposition 4.

We now consider binary random variables $X_1, \ldots, X_n$, each indicating the occurrence of a genetic event. Given a partially ordered set P, we define the discrete censored continuous-time conjunctive Bayesian network as the family of probability distributions

$$\text{pr}(X_1 = x_1, \ldots, X_n = x_n) = \text{pr}(T \vdash Q_S),$$

where $x_i \in \{0, 1\}$, $T \sim P_s$ and $S = \{i \in [n] \mid x_i = 1\}$. This model has state space $\{0, 1\}^n$, the set of all genotypes of length n, and parameters $\lambda = (\lambda_s, \lambda_1, \ldots, \lambda_n) \in \mathbb{R}^{n+1}_{>0}$. Although $n + 1$ parameters specify the model, it has dimension n, because λ can be rescaled.

Unlike in the fully observed continuous-time network or the discrete network, we have found no general closed-form expressions for the maximum likelihood estimates of the parameters of

the censored model. However, as the censored model is a marginalization of the continuous-time network, which is a regular exponential family, we can use the EM algorithm to find maximum likelihood estimates. While the EM algorithm is only guaranteed to find a local maximum of the likelihood function, our computational experience has been that using exact maximum likelihood estimates from the discrete conjunctive Bayesian network for θ_i and solving for λ_i in the approximate expression $\theta_i = \lambda_i/(\lambda_i + \lambda_s)$ works as a good starting guess.

In the EM algorithm, we start with a guess for the maximum likelihood parameters λ^* and compute the expected values of the sufficient statistics of the fully observed model, given the data. Specifically, for each observed $S \subseteq [n]$ and each $i \in [n]$, we compute the expected value

$$E\left(T_i - \max_{j \in \mathrm{pa}(i)} T_j \mid T \vdash Q_S \right). \tag{5}$$

This is the E-step of the algorithm. In the M-step, the expected sufficient statistics are used to compute maximum likelihood estimates for λ in the fully observed model. The EM algorithm iterates alternations of the E-step and the M-step. In each iteration, the likelihood function increases and a fixed point is a critical point of the likelihood function. The M-step is calculated by Proposition 1, while, in the E-step, we have the following formula for the expected value (5).

THEOREM 4.   *The expected value of $T_i - \max_{j \in \mathrm{pa}(i)} T_j$ given that T is compatible with Q is*

$$E\left(T_i - \max_{j \in \mathrm{pa}(i)} T_j \mid T \vdash Q \right) = \frac{\lambda_s \lambda_1 \cdots \lambda_n}{\mathrm{pr}(T \vdash Q)} \sum_{C \subset \mathcal{C}\{J(Q)\}} \left(\prod_{k=0}^{n} \frac{1}{\lambda_{\mathrm{Exit}(C_k)}} \right) \left\{ \sum_{l=0}^{n} \frac{\iota(i, C_l)}{\lambda_{\mathrm{Exit}(C_l)}} \right\},$$

where the first sum is over all maximal chains in $J(Q)$ and

$$\iota(i, C_k) = \begin{cases} 1, & i \notin C_k, \mathrm{pa}(i) \subseteq C_k, \\ 0, & otherwise. \end{cases}$$

Proof. The proof follows the pattern of the proof of Theorem 2. The expected value is

$$E\left(T_i - \max_{j \in \mathrm{pa}(i)} T_j \mid T \vdash Q \right) = \frac{1}{\mathrm{pr}(T \vdash Q)} \int_{\mathbb{R}_{\geqslant 0}^{n+1}} \left(t_i - \max_{j \in \mathrm{pa}(i)} t_j \right) \mathbb{I}_Q(t) f(t) dt.$$

We can calculate the integral by decomposing it into a sum over the linear extensions of Q, i.e. the chains in the distributive lattice $J(Q)$. Without loss of generality, we can suppose that the linear extension is called $0 \prec \cdots \prec n$. For this linear extension, the integral becomes

$$\int_{t_0=0}^{\infty} \int_{t_1=t_0}^{\infty} \cdots \int_{t_n=t_{n-1}}^{\infty} \left(t_i - t_{j(i)} \right) f(t) dt$$

and over this region $f(t) = \prod_{k=0}^{n} \lambda_k \exp\{-\lambda_k(t_k - t_{j(k)})\}$. Applying the usual change of coordinates, we can rewrite the integral in product form as

$$\int_{u_0=0}^{\infty} \cdots \int_{u_n=0}^{\infty} \left(u_{i-1} + \cdots + u_{j(i)} \right) \prod_{k=0}^{n} \lambda_k \exp\{-\lambda_k \left(u_{k-1} + u_{k-2} + \cdots + u_{j(k)} \right)\} du.$$

Breaking this integral up as a sum yields a collection of integrals we have already computed in the proof of Theorem 2. However, each term

$$\prod_{k=0}^{n} \frac{1}{\lambda_{\mathrm{Exit}(C_k)}} \frac{1}{\lambda_{\mathrm{Exit}(C_l)}}$$

contributes to the sum, if and only if $i \in \mathrm{Exit}(C_l)$, i.e. if and only if $\iota(i, C_l) = 1$.   □

Rather than computing the expectation from Theorem 4 by explicitly listing all maximal chains in the distributive lattice $J(Q)$, the expected value can be computed recursively, by summing up the distributive lattice. This dynamic programming approach reduces the computational burden of computing the expectation, because one need not enumerate all maximal chains in $J(Q)$.

PROPOSITION 4. *For each $S \in J(Q)$ define P_S and E_S^i by the formulae*

$$P_S = \sum_{j \in S : S \setminus \{j\} \in J(Q)} \frac{\lambda_j}{\lambda_{\mathrm{Exit}(S \setminus \{j\})}} P_{S \setminus \{j\}},$$

$$E_S^i = \sum_{j \in S : S \setminus \{j\} \in J(Q)} \left\{ \frac{\lambda_j}{\lambda_{\mathrm{Exit}(S \setminus \{j\})}} E_{S \setminus \{j\}}^i + \iota(i, S \setminus \{j\}) \frac{\lambda_j}{\lambda_{\mathrm{Exit}(S \setminus \{j\})}^2} P_{S \setminus \{j\}} \right\},$$

subject to the initial conditions $P_\emptyset = 1$ and $E_\emptyset^i = 0$, where

$$\iota(i, S \setminus \{j\}) = \begin{cases} 1, & i \notin S \setminus \{j\}, \, \mathrm{pa}(i) \subseteq S \setminus \{j\}, \\ 0, & \textit{otherwise}. \end{cases}$$

Then $\mathrm{pr}(T \vdash Q) = P_{\{s\} \cup [n]}$ and $E(T_i - \max_{j \in \mathrm{pa}(i)} T_j \mid T \vdash Q) = E_{\{s\} \cup [n]}^i / P_{\{s\} \cup [n]}$.

Proof. Both results follow from writing down a closed-form expression for P_S and E_S^i, proving that these formulae hold inductively, and showing that $P_{\{s\} \cup [n]} = \mathrm{pr}(T \vdash Q)$ and $E_{\{s\} \cup [n]}^i / P_{\{s\} \cup [n]} = E(T_i - \max_{j \in \mathrm{pa}(i)} T_j \mid T \vdash Q)$.

To this end, let $Q \mid_S$ be the induced sub-partially ordered set of Q with element set S. Then

$$P_S = \prod_{i \in S} \lambda_i \sum_{C \in \mathcal{C}\{J(Q_S)\}} \prod_{i=0}^{|S|-1} \frac{1}{\lambda_{\mathrm{Exit}(C_i)}}$$

with $P_\emptyset = 1$. The recurrence

$$P_S = \sum_{j \in S : S \setminus \{j\} \in J(Q)} \frac{\lambda_j}{\lambda_{\mathrm{Exit}(S \setminus \{j\})}} P_{S \setminus \{j\}}$$

is satisfied because every maximal chain in $J(Q \mid_S)$ comes from a maximal chain in exactly one of the $J(Q \mid_{S \setminus \{j\}})$ by adding j to $Q \mid_{S \setminus \{j\}}$ as the last element. Also, $P_{\{s\} \cup [n]}$ has the desired form. Similarly, it is straightforward to show that

$$E_S^i = \prod_{k \in S} \lambda_k \sum_{C \subset \mathcal{C}\{J(Q_S)\}} \left(\prod_{k=0}^{|S|-1} \frac{1}{\lambda_{\mathrm{Exit}(C_k)}} \right) \left\{ \sum_{k=0}^{|S|-1} \frac{\iota(i, C_k)}{\lambda_{\mathrm{Exit}(C_k)}} \right\},$$

which, together with P_S, above, satisfies the desired recurrence relation. $\square$

5. APPLICATIONS

5·1. *Noise model*

Theorem 1 states that the structure of the maximum likelihood continuous-time conjunctive Bayesian network is given by the maximal partially ordered set that is compatible with the data. In practice, however, the observations are subject to noise, either due to deviations of the data generating process from the model, or due to technical limitations in assessing genetic changes experimentally. Thus, for most biomedical datasets, the maximum likelihood partially ordered

set will have few relations, although many observations might support more order constraints. We address this problem using a simple mixture model (Beerenwinkel et al., 2007).

For the censored model (4) and observed data $u : \{0, 1\}^n \to \mathbb{N}_{\geqslant 0}$, where u_g denotes the count of genotype g in the data, we define a family of partially ordered sets P_ϵ, $0 \leqslant \epsilon \leqslant 1$, as follows. We consider all possible relations in ascending order of the number of observations that they violate. Starting with the partially ordered set P_ϵ with no relation, we add, in a greedy fashion, each relation $i \prec j$, if the number of its violations is smaller than $N\epsilon$ and if $j \prec i$ is not in P_ϵ.

We assume that incompatible genotypes are generated with uniform probability $q_\epsilon = 1/(2^n - |J(P_\epsilon)|)$ and consider the extended censored model with probabilities

$$\mathrm{pr}(T_\epsilon \vdash Q \mid \alpha, \lambda) = \begin{cases} \alpha \, \mathrm{pr}(T_\epsilon \vdash Q \mid \lambda), & \text{if } Q \text{ refines } P_\epsilon, \\ (1 - \alpha) \, q_\epsilon, & \text{otherwise,} \end{cases}$$

where $T_\epsilon \sim P_\epsilon$ and $\alpha = \sum_{g \in J(P_\epsilon)} u_g / \sum_{g \in 2^{[n]}} u_g$ denotes the fraction of data compatible with P_ϵ. The model can be regarded as a mixture model with α the maximum likelihood estimate of the mixing parameter.

In the applications, we construct partially ordered sets P_ϵ for various values of ϵ and select the one that maximizes the likelihood of the extended model. A software implementation of the model including the algorithms for model selection and parameter estimation is available at http://www.cbg.ethz.ch/software/ct-cbn.

5·2. *Cancer data*

We first use the extended censored model to analyze the genetic progression of prostate cancer. The random variables T_i denote the times of fixation of genetic changes in the population of cancer cells. Genetic changes were assessed by comparative genome hybridization experiments. This technique detects large-scale genomic alterations, namely the gain or loss of chromosome arms.

For example, the event 4q+ denotes the gain (+) of additional copies of the large (q) arm of chromosome 4. Likewise, 8p– refers to the loss (–) of the small arm (p) of chromosome 8. We consider 54 prostate cancer samples, each defined by the presence or absence of the nine alterations 3q+, 4q+, 6q+, 7q+, 8p–, 8q+, 10q–, 13q+ and Xq+ (Rahnenführer et al., 2005).

In Fig. 2(a), the loglikelihood is shown as a function of the fraction of incompatible genotypes. The partially ordered set that maximizes the likelihood explains 89% of the data and is displayed in Fig. 2(b). Four of the nine genetic changes do not obey any relation, two events have one direct predecessor and one event occurs only after two parent events have occurred. The second best partially ordered set is the one with no relations, corresponding to $\alpha = 1$.

In order to assess the uncertainty associated with parameter estimation and model selection, we performed two independent bootstrap analyses comprising 10 000 samples, each of size $N = \sum_g u_g = 54$. First, we fixed the optimal partially ordered set, Fig. 2(b), and re-estimated the model parameters λ_i. The quartiles of the distributions are displayed in Table 1. The stability of optimal partially ordered sets was assessed by repeated model selection with a fixed tolerance parameter of $\epsilon = 0.05$, the value that had given rise to the optimal partial order with $\alpha = 0.89$. Figure 3 summarizes separately the cover relations and all relations that appeared in the bootstrap partially ordered sets. This analysis supports exactly the relations encoded in the partially ordered set of Fig. 2(b), especially the sequence 4q+ $\prec$ 8q+ $\prec$ 13q+, and 3q+ as a late event.

5·3. *HIV data*

Our second application is concerned with the evolution of drug resistance in HIV. We study the accumulation of amino acid changes in the viral target protein reverse transcriptase. The seven
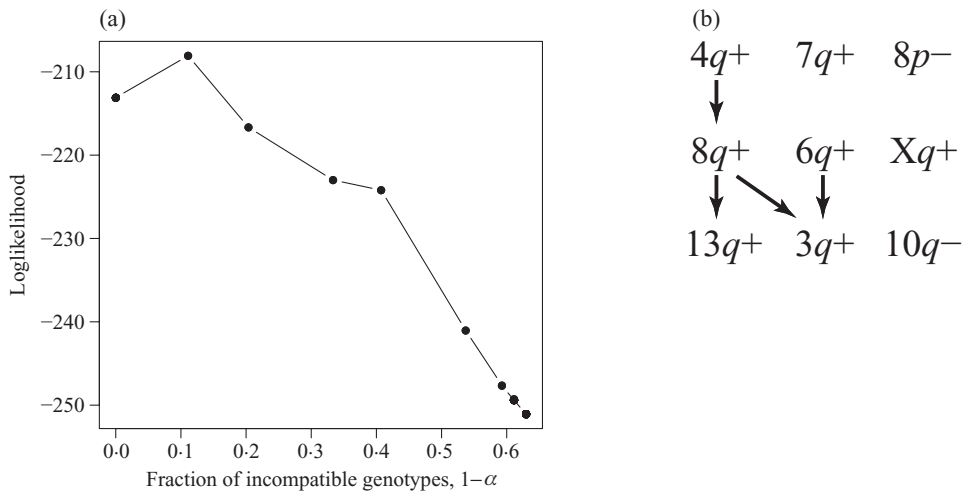
Fig. 2. Maximum likelihood estimation of the discrete censored continuous-time conjunctive Bayesian network model for the prostate cancer data. In (a), the loglikelihood is displayed as a function of the fraction of data that is incompatible with the partially ordered set. The curve has been generated by densely sampling ϵ from the unit interval and estimation of the extended models P_ϵ. The optimal prostate cancer partially ordered set corresponding to the maximum of the graph is displayed in panel (b). An arrow $p \rightarrow q$ between two genetic events represents the cover relation $p \prec q$ in the Hasse diagram.

Table 1. *Bootstrap analysis of model parameters for the optimal prostate cancer partially ordered set of Fig. 2. The partially ordered set structure is repeated by listing the parents of each genetic event, i.e. chromosomal alteration. The quartiles of the distribution of each parameter λ_i are given in columns Q1, 25% quantiles; Q2, 50% quantiles, median; and Q3, 75% quantiles*

Event	Parents	Q1	Q2	Q3
10q–		0·09	0·13	0·17
8p–		0·17	0·21	0·27
13q+	8q+	1·17	1·82	2·92
3q+	6q+, 8q+	0·95	1·68	2·98
4q+		0·27	0·35	0·44
6q+		0·27	0·35	0·43
7q+		0·15	0·20	0·26
8q+	4q+	2·20	3·35	6·45
Xq+		0·26	0·33	0·41

resistance-associated mutations 41L, 67N, 69D, 70R, 210W, 215Y and 219Q (Johnson et al., 2008) are considered, where, for example, 41L indicates the presence of the amino acid leucine at position 41 of the reverse transcriptase. A total of 364 viruses are analyzed that have been isolated from infected patients under antiretroviral therapy. Here the random variables T_i denote the fixation times of the mutations in the virus population.

The optimal partially ordered set for the HIV drug resistance data explains 87% of the observations, Fig. 4(a). Its Hasse diagram has two connected components, Fig. 4(b). The first one represents the linear pathway 215Y $\prec$ 41L $\prec$ 210W, and in the second one, mutations 70R, 67N, 219Q and 69D, form a rhombus beginning with 70R and ending with 69D. This finding supports

(a) Cover relations

	10q−	8p−	13q+	3q+	4q+	6q+	7q+	8q+	Xq+
10q−		0.5	0	0.8			0.1		
8p−	4.9		0.1	0.7	0	0	2.1	0.1	0.1
13q+	2.7	0.4		28.8	3	3.5	6.9	15.8	0.7
3q+	4.7	0.4	3		0.6	0.5	1.9	1.6	0.1
4q+	4.7	1.6	11.7	22.3		17.6	4.7	55.4	1
6q+	17.6	0.8	28.7	42.4	15.7		13.6	13.2	1.8
7q+	4.3	4.7	3.3	7.8	0.6	0.5		4.2	0.4
8q+	0.6	3	51.8	39.2	16.8	6.6	32.4		1.6
Xq+	4.2	4.8	4.4	9.4	0.9	0.8	4.8	1.8	

(b) All relations

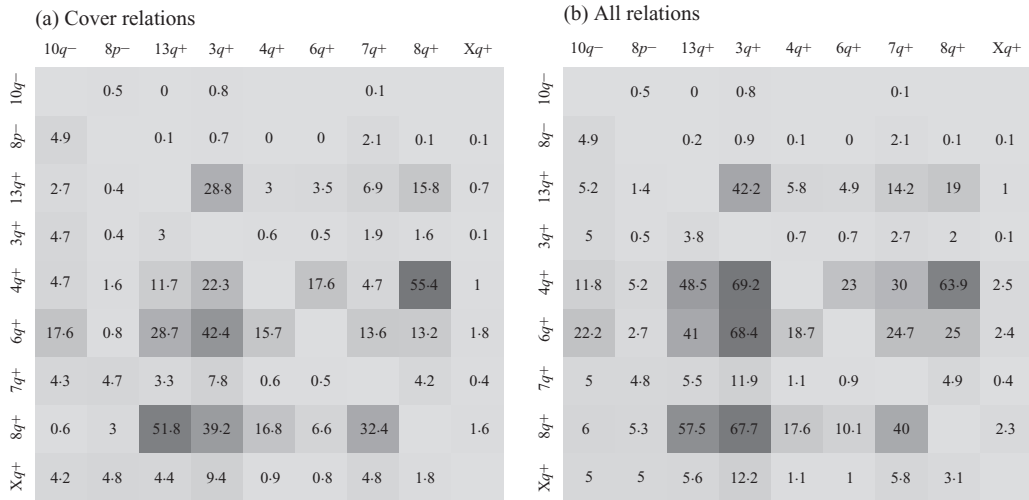|	10q−	8p−	13q+	3q+	4q+	6q+	7q+	8q+	Xq+
10q−		0.5	0	0.8			0.1		
8p−	4.9		0.2	0.9	0.1	0	2.1	0.1	0.1
13q+	5.2	1.4		42.2	5.8	4.9	14.2	19	1
3q+	5	0.5	3.8		0.7	0.7	2.7	2	0.1
4q+	11.8	5.2	48.5	69.2		23	30	63.9	2.5
6q+	22.2	2.7	41	68.4	18.7		24.7	25	2.4
7q+	5	4.8	5.5	11.9	1.1	0.9		4.9	0.4
8q+	6	5.3	57.5	67.7	17.6	10.1	40		2.3
Xq+	5	5	5.6	12.2	1.1	1	5.8	3.1	

Fig. 3. Bootstrap analysis of partially ordered set structures for the prostate cancer comparative genome hybridization data. The entry and the degree of shading in matrix cell (i, j) denote the relative frequency in percent of the relation $i \prec j$ among the bootstrap samples. In panel (a) only cover relations are recorded, whereas in panel (b) all relations are considered. Diagonal entries and entries identical to zero have been omitted.
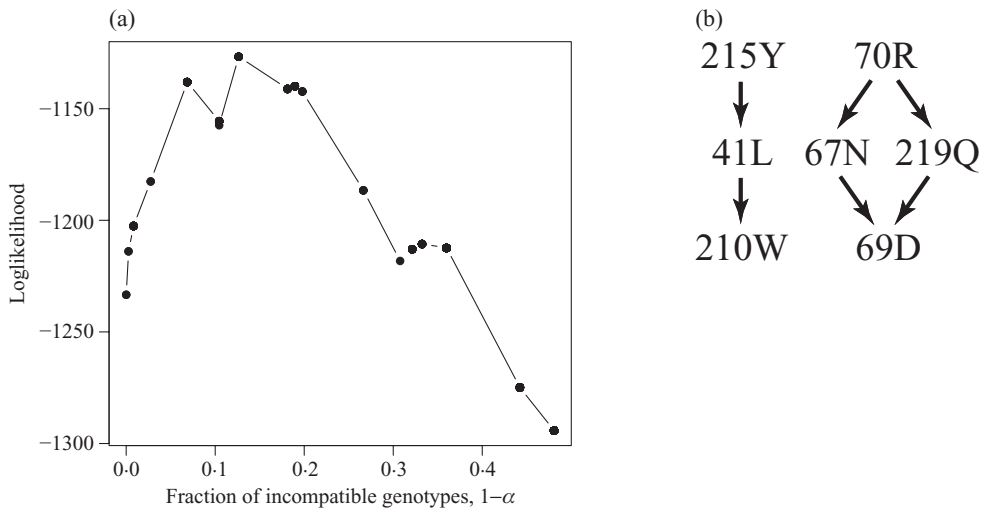

Fig. 4. Maximum likelihood estimation of the continuous-time conjunctive Bayesian network model for the HIV drug resistance data (a), and optimal HIV drug resistance partially ordered set corresponding to the maximum of the graph (b).

previous studies in which the same clustering of mutations has been described (Boucher et al., 1992). The two groups of mutations are known as the '215-41 pathway' and the '70-219 pathway', respectively. They provide alternative but not exclusive routes to resistance for HIV. The continuous-time conjunctive Bayesian network model captures this escape behaviour and suggests order constraints within each group.

A similar bootstrap analysis was performed for the HIV data and the results are summarized in Table 2 and Fig. 5. We find high stability of the maximum likelihood partially ordered set under re-sampling. All cover relations of Fig. 4 are supported with high confidence, and additional relations reach only less than half of these values. Overall, these findings strengthen the picture

Table 2. *Bootstrap analysis of model parameters for the optimal HIV partially ordered set of Fig. 4. Events are amino acid changes in the HIV reverse transcriptase. See caption of Table 1 for further details*

Event	Parents	Q1	Q2	Q3
41L	215Y	1·78	2·06	2·36
67N	70R	0·92	1·03	1·16
69D	67N, 219Q	0·55	0·69	0·86
70R		0·59	0·62	0·67
210W	41L	0·61	0·72	0·85
215Y		0·52	0·55	0·59
219Q	70R	0·78	0·88	1·00
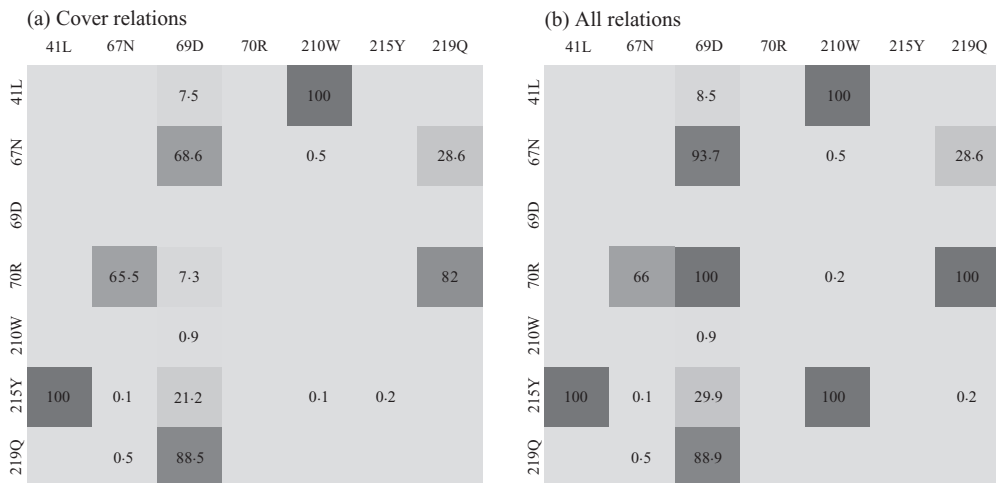
Fig. 5. Bootstrap analysis of partially ordered set structures for the HIV mutation data. See caption of Fig. 3 for further details.

of two separate groups of mutations, one occurring in a linear order, $215Y \prec 41L \prec 210W$, and the other exhibiting two intermediate branches, $70R \prec (67N, 219Q) \prec 69D$.

For both applications, the best tree model shows inferior performance as compared to the more general model based on partially ordered sets. For example, the mutagenetic tree model for prostate cancer found in Fig. 2 of Rahnenführer et al. (2005) explains only 56% of the data at a loglikelihood of only $-248\cdot4$. Unlike mutagenetic trees, continuous-time conjunctive Bayesian networks can model the requirement of multiple parent mutations. This type of order constraint was found in both the prostate cancer and the HIV data set.

6. DISCUSSION

Conjunctive Bayesian networks are statistical models for the accumulation of mutations. They are defined by a partially ordered set of mutations, which encodes constraints on the order in which mutations can occur. Here we have introduced the continuous-time conjunctive Bayesian network, a continuous-time version of this model in which each mutation appears after an exponentially distributed waiting time, provided that all predecessor mutations have occurred. In an evolutionary process, this waiting time includes the generation of the mutation plus the time it takes for the allele to reach fixation in the population. Since we consider only mutations

with a selective advantage, the waiting time will be dominated by the mutation process in large populations and by the fixation process in small populations. The parameters λ correspond to the rate of evolution, i.e. the product of population size, mutation rate and fixation probability.

The continuous-time conjunctive Bayesian network gives information on the order in which mutations tend to occur. For a set of n mutations, the number of possible pathways to evolve the wild type into the type harbouring all n mutations is the number of linear extensions of the partially ordered set. This cardinality increases rapidly with n (Brightwell & Winkler, 1991). However, protein evolution appears to use few mutational pathways (Weinreich et al., 2006). Thus, we can expect to find partially ordered sets with much smaller lattices of order ideals than the full Boolean lattice. In this reduced genotype space, evolution can be modelled more efficiently.

Knowledge of the partial order of genetic events helps our understanding of the phenotypic changes and biological mechanisms driving the evolutionary process. Furthermore, it allows for identifying early and essential mutational steps that may be predictive of clinical outcome or point to promising drug targets. In cancer research, Fearon & Vogelstein (1990) have proposed linear pathways of genetic alterations as a model of tumorigenesis. These models are known as Vogelgrams (Gatenby & Maini, 2003; Jones et al., 2008). The continuous-time conjunctive Bayesian network can be regarded as a generalization of the Vogelgram that is equipped with a statistical methodology for model selection and parameter estimation. In particular, it allows for multiple evolutionary pathways and makes explicit the timeline for the genetic alterations.

We have derived equations for the maximum likelihood estimates of the model parameters and for the expected waiting time of any genotype. These results are used in the EM algorithm for the censored model. Censoring is modelled by an exponentially distributed sampling time of the observed genotypes. This model appears most relevant for the data sets available, which often comprise cross-sectional data sampled after unknown time periods with respect to the evolutionary process. Other censoring schemes might be applicable in the future. For example, the sampling time, but not the time of appearance of each mutation, can be observed in some situations, giving rise to a different marginalization of the continuous-time network.

Because model selection relies on a simple combinatorial criterion and the number of model parameters is only linear in the number of mutations, we expect the continuous-time conjunctive Bayesian network to scale well with increasing data sets both in the number of observations and the number of mutations. In the cancer and HIV applications presented here, there are between 7 and 12 genetic events and 35–364 observations. It is likely, however, that the number of genes associated with cancer progression is much larger than currently known. Because the running time of the EM algorithm is dominated by the size of the genotype lattice, many mutations can be modelled as long as the number of mutational pathways is limited.

REFERENCES

BEERENWINKEL, N. & DRTON, M. (2007). A mutagenetic tree hidden Markov model for longitudinal clonal HIV sequence data. *Biostatistics* **8**, 53–71.

BEERENWINKEL, N., DÄUMER, M., SING, T., RAHNENFÜHRER, J., LENGAUER, T., SELBIG, J., HOFFMANN, D. & KAISER, R. (2005a). Estimating HIV evolutionary pathways and the genetic barrier to drug resistance. *J. Inf. Dis.* **191**, 1953–60.

BEERENWINKEL, N., ERIKSSON, N. & STURMFELS, B. (2006). Evolution on distributive lattices. *J. Theor. Biol.* **242**, 409–20.

BEERENWINKEL, N., ERIKSSON, N. & STURMFELS, B. (2007). Conjunctive Bayesian networks. *Bernoulli* **13**, 893–909.

BEERENWINKEL, N., RAHNENFÜHRER, J., DÄUMER, M., HOFFMANN, D., KAISER, R., SELBIG, J. & LENGAUER, T. (2005b). Learning multiple evolutionary pathways from cross-sectional data. *J. Comp. Biol.* **12**, 584–98. RECOMB 2004.

BEERENWINKEL, N., RAHNENFÜHRER, J., KAISER, R., HOFFMANN, D., SELBIG, J. & LENGAUER, T. (2005c). Mtreemix: a software package for learning and using mixture models of mutagenetic trees. *Bioinformatics* **21**, 2106–07.

BOUCHER, C. A., O'SULLIVAN, E., MULDER, J. W., RAMAUTARSING, C., KELLAM, P., DARBY, G., LANGE, J. M., GOUDSMIT, J. & LARDER, B. A. (1992). Ordered appearance of zidovudine resistance mutations during treatment of 18 human immunodeficiency virus-positive subjects. *J. Inf. Dis.* **165**, 105–10.

BRIGHTWELL, G. R. & WINKLER, P. (1991). Counting linear extensions. *Order* **8**, 225–42.

DEFORCHE, K., SILANDER, T., CAMACHO, R., GROSSMAN, Z., SOARES, M. A., LAETHEM, K. V., KANTOR, R., MOREAU, Y., VANDAMME, A.-M. & NON B. WORKGROUP (2006). Analysis of HIV-1 pol sequences using Bayesian networks: implications for drug resistance. *Bioinformatics* **22**, 2975–79.

DESPER, R., JIANG, F., KALLIONIEMI, O. P., MOCH, H., PAPADIMITRIOU, C. H. & SCHÄFFER, A. A. (1999). Inferring tree models for oncogenesis from comparative genome hybridization data. *J. Comp. Biol.* **6**, 37–51.

FEARON, E. R. & VOGELSTEIN, B. (1990). A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767.

FOULKES, A. & DEGRUTTOLA, V. (2003). Characterizing the progression of viral mutations over time. *J. Am. Statist. Assoc.* **98**, 859–67.

GATENBY, R. A. & MAINI, P. K. (2003). Mathematical oncology: cancer summed up. *Nature* **421**, 321.

HJELM, M., HÖGLUND, M. & LAGERGREN, J. (2006). New probabilistic network models and algorithms for oncogenesis. *J. Comp. Biol.* **13**, 853–65.

IWASA, Y., MICHOR, F. & NOWAK, M. A. (2003). Evolutionary dynamics of escape from biomedical intervention. *Proc. Biol. Sci.* **270**, 2573–78.

JOHNSON, V. A., BRUN-VEZINET, F., CLOTET, B., GUNTHARD, H. F., KURITZKES, D. R., PILLAY, D., SCHAPIRO, J. M. & RICHMAN, D. D. (2008). Update of the drug resistance mutations in hiv-1: Spring 2008. *Top. HIV Med.* **16**, 62–68.

JONES, S., CHEN, W.-D., PARMIGIANI, G., DIEHL, F., BEERENWINKEL, N., ANTAL, T., TRAULSEN, A., NOWAK, M. A., SIEGEL, C., VELCULESCU, V. E., KINZLER, K. W., VOGELSTEIN, B., WILLIS, J. & MARKOWITZ, S. D. (2008). Comparative lesion sequencing provides insights into tumor evolution. *Proc. Nat. Acad. Sci.* **105**, 4283–88.

NORRIS, J. (1997). *Markov Chains*. Cambridge, UK: Cambridge University Press.

RADMACHER, M. D., SIMON, R., DESPER, R., TAETLE, R., SCHÄFFER, A. A. & NELSON, M. A. (2001). Graph models of oncogenesis with an application to melanoma. *J. Theor. Biol.* **212**, 535–48.

RAHNENFÜHRER, J., BEERENWINKEL, N., SCHULZ, W. A., HARTMANN, C., VON DEIMLING, A., WULLICH, B. & LENGAUER, T. (2005). Estimating cancer survival and clinical outcome based on genetic tumor progression scores. *Bioinformatics* **21**, 2438–46.

SIMON, R., DESPER, R., PAPADIMITRIOU, C. H., PENG, A., ALBERTS, D. S., TAETLE, R., TRENT, J. M. & SCHÄFFER, A. A. (2000). Chromosome abnormalities in ovarian adenocarcinoma: Iii. using breakpoint data to infer and test mathematical models for oncogenesis. *Genes Chromosomes Cancer* **28**, 106–20.

STANLEY, R. (1999). *Enumerative Combinatorics*. Cambridge, UK: Cambridge University Press.

VON HEYDEBRECK, A., GUNAWAN, B. & FÜZESI, L. (2004). Maximum likelihood estimation of oncogenetic tree models. *Biostatistics* **5**, 545–56.

WEINREICH, D. M., DELANEY, N. F., DEPRISTO, M. A. & HARTL, D. L. (2006). Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–14.